

QY 1 AGTGCCTGGGGCCCTCGGGGCTAGTACACAGCACCTGAGTGGSCACAGAGGAC 60
DB 1 AGTGCCTGGGGCCCTCGGGGCTAGTACACAGCACCTGAGTGGSCACAGAGGAC 60
QY 61 CCTCTCATGTTTAGGGAACCTCTGGGCTCAGAGAGCTGGGGCCCTCGGGGAG 120
DB 61 CCTCTCATGTTTAGGGAACCTCTGGGCTCAGAGAGCTGGGGCCCTCGGGGAG 120
QY 121 TCCGCCCATCCGGGGCGGAAATGCTCCGGGCTCGAGTGGCTGCTGGCTGCTCCC 180
DB 121 TCCGCCCATCCGGGGCGGAAATGCTCCGGGCTCGAGTGGCTGCTGGCTGCTCCC 180
QY 181 TGGTTGCTGGGTGCAAGTCTGGGTTCTGGGTTCTGGATTCGGGGCCGTTTACAGT 240
DB 181 TGGTTGCTGGGTGCAAGTCTGGGTTCTGGGTTCTGGATTCGGGGCCGTTTACAGT 240
QY 241 AGCTGTGGCGGCTCCTCGGCTGAGTCCGTCGGCGCGGTGCTCCGGGACGGCCTAGC 300
DB 241 AGCTGTGGCGGCTCCTCCT----- 257
QY 301 TGCCGGGGTCCGGGGCCCCAGGCAATTCGGGGCTGCAGATTGACGGGATTCGGGATGCA 360
DB 258 -----CGGATGCA 265
QY 361 CCGGGGCCCGGGCCCTCACGAGCGGTCCAGACCTGGTGGGAAGAGGTGGGGGAC 420
DB 266 CCGGGGCCCGGGCCCGCCCCACCGAGCGGTCCAGACCTGGTGGGAAGAGGTGGGGGAC 325
QY 421 GGGTCCCTGAGGATCCGATGCTCAGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA 480
DB 326 GGGTCCCTGAGGATCCGATGCTCAGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA 385
QY 481 CACATGTGATTTACCTCAGTTTGTGATCCGTAAATGGAACAATTCGAAGTACTTCA 540
DB 386 CACATGTGATTTACCTCAGTTTGTGATCCGTAAATGGAACAATTCGAAGTACTTCA 445
QY 541 CAGTGTGTTGAGAGGATTAATGAACAATGCTTGAAGCTCTTTGCAAGAGGAGGCC 600
DB 446 CAGTGTGTTGAGAGGATTAATGAACAATGCTTGAAGCTCTTTGCAAGAGGAGGCC 505
QY 601 TCGGAAGCAGGGCTGGCGGCGAGACACACCTGCTGTCCAGGAGCAACAGGAGCAT 660
DB 506 TCGGAAGCAGGGCTGGCGGCGAGACACACCTGCTGTCCAGGAGCAACAGGAGCAT 565
QY 661 GAAGACCCCGTGGAGTGGCGCTCAGTGGGATGACAGCCCTGGGCTTCAGCACCGCTG 720
DB 566 GAAGACCCCGTGGAGTGGCGCTCAGTGGGATGACAGCCCTGGGCTTCAGCACCGCTG 625
QY 721 CCGAGGTGGCTACCGGCTCAAGGCCAGGACGTATATGTGGATCAGACTCTGTTGGCAG 780
DB 626 CCGAGGTGGCTACCGGCTCAAGGCCAGGACGTATATGTGGATCAGACTCTGTTGGCAG 685
QY 781 CCCAGCAGGACCCGGGCTACCCACCGGACTTCGATCCGCGCTGGGTGGAGAGGCTAA 840
DB 686 CCCAGCAGGACCCGGGCTACCCACCGGACTTCGATCCGCGCTGGGTGGAGAGGCTAA 745
QY 841 CAGAACACAGAGGCTGGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGGAGCTGTGA 900
DB 746 CAGAACACAGAGGCTGGGCAAGGAGGATCGAAGTCTTGGGGCAAGGGGAGCTGTGA 805
QY 901 GACCACCCCTCAAGGGGCGACACCCCAACCTCACACCAAGGAAGAACAAATACAG 960
DB 806 GACCACCCCTCAAGGGGCGACACCCCAACCTCACACCAAGGAAGAACAAATACAG 865
QY 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTCCGCTGTTGGCTCCCGATCGAAGG 1020
DB 866 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTCCGCTGTTGGCTCCCGATCGAAGG 925
QY 1021 CGCCAGCTTCGGGGCCCGCGGATGCGGAGGGGATGCGGCAAGCTCCGCTGCTCTT 1080
DB 926 CGCCAGCTTCGGGGCCCGCGGATGCGGAGGGGATGCGGCAAGCTCCGCTGCTCTT 985

QY 1081 GTGACCGCACCACTACCCCGAGGGTAGCCACTCGCCCGGCCCGAGGAGGACCACT 1140
DB 986 GTGACCGCACCACTACCCCGAGGGTAGCCACTCGCCCGGCCCGAGGAGGACCACT 1045
QY 1141 GCGAGCCATTACCCAGTGGTCCCTCAAGACAGAGCCGGGGCCAGCGGAGACTCCCA 1200
DB 1046 GCGAGCCATTACCCAGTGGTCCCTCAAGACAGAGCCGGGGCCAGCGGAGACTCCCA 1105
QY 1201 GAAGTTATCTATGGTGGGTGTTACACTCTTCAGCGCCCTGAAGCGGGGACTTTCCCATTC 1260
DB 1106 GAAGTTATCTATGGTGGGTGTTACACTCTTCAGCGCCCTGAAGCGGGGACTTTCCCATTC 1165
QY 1261 CCTCACCACTGAATGTCCCGACACTGGTCAATCCAGCCACAGTGCCTCCCAACAAA 1320
DB 1166 CCTCACCACTGAATGTCCCGACACTGGTCAATCCAGCCACAGTGCCTCCCAACAAA 1225
QY 1321 TGGGCTCAGGATCTCAGGCTTCCAGTCAAGGGTGAACCTTCGGAGCCCTCTGGTGAC 1380
DB 1226 TGGGCTCAGGATCTCAGGCTTCCAGTCAAGGGTGAACCTTCGGAGCCCTCTGGTGAC 1285
QY 1381 TTCCAGGCTCGCTCAGTTCAGTGCATTTCCAGTGCATCTACCCACAGAGTGGGGCCAC 1440
DB 1286 TTCCAGGCTCGCTCAGTTCAGTGCATTTCCAGTGCATCTACCCACAGAGTGGGGCCAC 1345
QY 1441 CCAGAAACCAAGCCCTTGGAAATGATACTTTTCATCAGGTTGCCTATGGGGCCAC 1500
DB 1346 CCAGAAACCAAGCCCTTGGAAATGATACTTTTCATCAGGTTGCCTATGGGGCCAC 1405
QY 1501 GCGACAGGATGGCCCTTCCAGGTTAGAGAGACATTCATCACCAGGAAACCCAGG 1560
DB 1406 GCGACAGGATGGCCCTTCCAGGTTAGAGAGACATTCATCACCAGGAAACCCAGG 1465
QY 1561 TATTAAGAAAGCCCTGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTT 1620
DB 1466 TATTAAGAAAGCCCTGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTT 1525
QY 1621 ATCTGCAATCTCTAGTGCATTTCTGCTTTTCTCCGATTCGGATTCGGGGCCAC 1680
DB 1526 ATCTGCAATCTCTAGTGCATTTCTGCTTTTCTCCGATTCGGATTCGGGGCCAC 1585
QY 1681 CTCATAGATGCTCTCCAGCCCTGTCTCAACATCTCCAAATAGTGCACACCCAGG 1740
DB 1586 CTCATAGATGCTCTCTCCAGCCCTGTCTCAACATCTCCAAATAGTGCACACCCAGG 1645
QY 1741 GGCCTGGCACCTCCACATCATCTCTTGTGTCGAAGTGCAGATTAACGGGCTGAT 1800
DB 1646 GGCCTGGCACCTCCACATCATCTCTTGTGTCGAAGTGCAGATTAACGGGCTGAT 1705
QY 1801 TGCC 1804
DB 1706 TGCC 1709

RESULT 8
AX880429
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.,
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 15334 07-FEB-2001;
Research Association for Biotechnology (JP)
Location/Qualifiers
1. .1614
/organism="Homo sapiens"

```
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
635..1168
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE90944.1"
/db_xref="GI:40035166"
/translaton="MKTPVELAVSMQTLGHRGGRVYKARTSVDETLPFGSPAG
TRPFPDFDPWFKARVGVKASKALGAGSCETTPSRGSTPLPRKKNKVRP
ISHTPSYCDLSFSGSESGASFGAPRVAKGDAKRALLTWPPTPRGSHSPRPREP
VPPQMGRLRISGLPRQG"

ORIGIN
      79 24; Score 1434.8; DB 6; Length 1614;
Query Match
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
QY 25 AGTACACAGCAGCTGAGTGGCCACAGAGACCTCTCCATGTTTAGGACCTCCT 84
DB 1 AGTACACAGCAGCTGAGTGGCCACAGAGACCTCTCCATGTTTAGGACCTCCT 60
QY 85 GGGCTCAGAGCTGGGCGCCCGCTGGGCGGACTCCCCCATCCGGGGCGCAATG 144
DB 61 GGGCTCAGAGCTGGGCGCCCGCTGGGCGGACTCCCCCATCCGGGGCGGATG 120
QY 145 GTCCGGGTCCGGTCCGAGTGTCTGGTGTCTCCCTGGTGTGGTGCAAAAGTCTG 204
DB 121 GTCCGGGGCGGTCCGAGTGTCTGGTGTCTCCCTGGTGTGGTGCAAAAGTCTG 180
QY 205 GTTCGTGGTCTTGGAATCCGGGCGCTTCAACGTAGCTGTCCGGCTCTCCGGGTGA 264
DB 181 GTTCGTGGTCTTGGAATCCGGGCGCTTCAACGTAGCTGTCCGGCTCTCCGGGTGA 240
QY 265 GTCCGTCCGGCGGGTCCCGGAGCGCTAGGCTGCGGGGTCGGGGGCGCCAGGC 324
DB 241 GTCCGTCCGGCGGGTCCCGGAGCGCTAGGCTGCGGGGTCGGGGGCGCCAGGC 300
QY 325 ATTCCGGGCTGCAGATTGACGGGATCCCGGATGCACCGCGCCCGCTCACCG 384
DB 301 ATTCCGGGCTGCAGATTGACGGGATCCCGGATGCACCGCGCCCGCTCACCG 360
QY 385 ACGGCTCCAGACCTGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCGATGCT 444
DB 361 ACGGCTCCAGACCTGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCGATGCT 420
QY 445 ACAGCCAGATGCTCAGCTTTATAGTGTGACTACACATGTGACTTCACCTCAGTTT 504
DB 421 ACAGCCAGATGCTCAGCTTTATAGTGTGACTACACATGTGACTTCACCTCAGTTT 480
QY 505 GTGATCCGTAATGGAATAATCGAAGTACTTCAAGTGTGTGAGGATTAATG 564
DB 481 GTGATCCGTAATGGAATAATCGAAGTACTTCAAGTGTGTGAGGATTAATG 540
QY 565 AAACAATGCTGTAAAGCTCTTTCAGGAGGAGCTCCGGAAGCGGCTTGGCGGCGAG 624
DB 541 AAACAATGCTGTAAAGCTCTTTCAGGAGGAGCTCCGGAAGCGGCTTGGCGGCGAG 600
QY 625 AGCACCTCTGTACACGGGACCAAGCAGCATGAAGACCCCGTGGAGCTGGCCGT 684
DB 601 AGCACCTCTGTACACGGGACCAAGCAGCATGAAGACCCCGTGGAGCTGGCCGT 660
QY 685 CAGTGGGATGCAGACCTCGGCTTCAGCACCGCTCGGAGTGGTACCGGTCAGGC 744
DB 661 CAGTGGGATGCAGACCTCGGCTTCAGCACCGCTCGGAGTGGTACCGGTCAGGC 720
QY 745 CAGGAGCTCATATGTGATGAGACTCTGTTTGGACGCCAGGACACCCGCTACCC 804
DB 721 CAGGAGCTCATATGTGATGAGACTCTGTTTGGACGCCAGGACACCCGCTACCC 780
QY 805 ACCGAGCTTCGATCCGCTTGGTGGAGAGGCTTACAGAACAGAGGCTGGGCAAGA 864
DB 781 ACCGAGCTTCGATCCGCTTGGTGGAGAGGCTTACAGAACAGAGGCTGGGCAAGA 840
```

```
QY 865 GGCATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAC 924
DB 841 GGCATCGAAGGCTTTGGGGGCAAGGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAC 900
QY 925 CCCCACCTCTACACCAAGGAAGAAACAATACAGACCCATCAGCCACACCCCTCTTA 984
DB 901 CCCCACCTCTACACCAAGGAAGAAACAATACAGACCCATCAGCCACACCCCTCTTA 960
QY 985 CTGTGATCAGTGGCTTTGGCTCCCGATCTGAAGGGGCCAGCTTCGGGGCCCGCGAT 1044
DB 961 CTGTGATCAGTGGCTTTGGCTCCCGATCTGAAGGGGCCAGCTTCGGGGCCCGCGAT 1020
QY 1045 GCGAAGGGGATCGCCAAAGCTCCGTCTCTCTTGTGGAGCGCACCACTACCCAG 1104
DB 1021 GCGAAGGGGATCGCCAAAGCTCCGTCTCTCTTGTGGAGCGCACCACTACCCAG 1080
QY 1105 GGTAGCCACTCGGCCCGCCCGGAGGACCACTCGAGCCATTACCCAGCTGGTCC 1164
DB 1081 GGTAGCCACTCGGCCCGCCCGGAGGAGG----- 1109
QY 1185 CTCGAAGACAGAGCGGGGCCAGCGGAGACTCCAGAAAGTTATCTATGGTGGGTACA 1224
DB 1110 ----- 1109
QY 1225 CTCCTACGCCCCCTGAAGCGGGACTTTCCCAATTCCTCACCCACCTGAATGTCCCAG 1284
DB 1110 ----- 1109
QY 1285 CACTGGTCACTCCAGCCACAGTCCCCCACAACAAATGGGCTCAGATCTCAGGCTTC 1344
DB 1110 -----CACCAGTCCCCCACAACAAATGGGCTCAGGATCTCAGGCTTC 1154
QY 1345 CACGTACGGGTGACCTTCGGGAGCCCTGGTACTTCCAGGGCTCGCTCAGTTAGCAT 1404
DB 1155 CACGTACGGGTGACCTTCGGGAGCCCTGGTACTTCCAGGGCTCGCTCAGTTAGCAT 1214
QY 1405 TTCACTGTCATCTACCCCAACAGAGTGGGGCCACCCAGAAACCAAGGCCCTTTGAA 1464
DB 1215 TTCACTGTCATCTACCCCAACAGAGTGGGGCCACCCAGAAACCAAGGCCCTTTGAA 1274
QY 1465 ATGATATCTTTTCATCAGGTTGCTATGGGGCCACCGGACAGGATAGCCCTTGCCA 1524
DB 1275 ATGATATCTTTTCATCAGGTTGCTATGGGGCCACCGGACAGGATAGCCCTTGCCA 1334
QY 1525 GGTAGAGAGACATTCATCACCAGGAAACCCAGGATTTAAAGAGGCCCTGTGGGGC 1584
DB 1335 GGTAGAGAGACATTCATCACCAGGAAACCCAGGATTTAAAGAGGCCCTGTGGGGC 1394
QY 1585 AGACAGATAGCAGGGGTGGGAGTGCCTCCCTTTATCCCTGACAAATCTCTAGTCGATTC 1644
DB 1395 AGACAGATAGCAGGGGTGGGAGTGCCTCCCTTTATCCCTGACAAATCTCTAGTCGATTC 1454
QY 1645 TTGCTTTTCTCCGATTTGGGATTTGGGGCCACCTCTAAGATGCTCTCCAGCCC 1704
DB 1455 TTGCTTTTCTCCGATTTGGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1514
QY 1705 TGCTCAACCATATCTCAAAATTAGTCCAAACCCAGGGGCTGGCACTCCCATCATATCC 1764
DB 1515 TGCTCAACCATATCTCAAAATTAGTCCAAACCCAGGGGCTGGCACTCCCATCATATCC 1574
QY 1765 ATTGCTTGTGTCAGGTCGAATTAACCGGCTGATTGCC 1804
DB 1575 ATTGCTTGTGTCAGGTCGAATTAACCGGCTGATTGCC 1614
```

RESULT 9
BD158376
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

BD158376 1614 bp DNA linear PAT 17-JAN-2003
Primer for synthesizing full-length cDNA and use thereof.
BD158376
BD158376.1 GI:27864134
JP 2002191363-A/13219.
Homo sapiens (human)

ORGANISM	Homo sapiens	Db
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	Qy
AUTHORS	1 (bases 1 to 1614)	Db
TITLE	Ota T., Isogai T., Nishikawa T., Hayashi K., Saito K., Yamamoto J., Ishii S., Sugiyama T., Wakanatsu A., Nagai K. and Otsuki T.	Qy
JOURNAL	Primer for synthesizing full-length cDNA and use thereof	Db
COMMENT	Patent: JP 2002191363-A 13219 09-JUL-2002; HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002191363-A/13219 PD 09-JUL-2002 PF 28-JUL-2000 JP 2002082090 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAOJU SUGIYAMA, T. WAKANATSU, A. NAGAI, K. SAITO, PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N5/00, C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers FT CDS Location/Qualifiers source 1..1614 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"	Qy
FEATURES	Query Match 79.2%; Score 1434.8; DB 6; Length 1614; Best Local Similarity 90.6%; Pred. No. 0; Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1; 1 AGTACACACGACCTGAGTGGAGTGGACACGAGAGCCCTTCCATGTTTAGGACCTCT 84 1 AGTACACACGACCTGAGTGGAGTGGACACGAGAGCCCTTCCATGTTTAGGACCTCT 60 85 GGGCTCAGAGAGTGGCGCGCGCCCTGGCGGACCTCCCGGATCGCGGGCGGAATG 144 61 GGGCTCAGAGAGTGGCGCGCGCCCTGGCGGACCTCCCGGATCGCGGGCGGAATG 120 145 GTCCGGTGGTCCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 204 121 GTCCGGTGGTCCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180 205 GTTCGGTGGTTCGAGTTCGGGGCGCTTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 264 181 GTTCGGTGGTTCGAGTTCGGGGCGCTTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240 265 GTCCGTCCGGCGGGTCCCGGGAGCGGCTAGGCTCCCGGGTCCCGGGCGCCAGGC 324 241 GTCCGTCCGGCGGGTCCCGGGAGCGGCTAGGCTCCCGGGTCCCGGGCGCCAGGC 300 325 ATTCGGGGTGCAGATTGACGGGATCCCGGATGCAACCGCGCGCGCGCGCGCGCGCGCG 384 301 ATTCGGGGTGCAGATTGACGGGATCCCGGATGCAACCGCGCGCGCGCGCGCGCGCGCG 360 385 ACGGGTCCAGACCTGGTGGAGAGTGGGGAGCGGTCCCTGAGGATCCCGATGCGCT 444 361 ACGGGTCCAGACCTGGTGGAGAGTGGGGAGCGGTCCCTGAGGATCCCGATGCGCT 420 445 ACAGGCCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTACCTCAGTTTT 504 421 ACAGGCCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTACCTCAGTTTT 480 505 GTGATCCGTAATAATGGACAAATTCGAGTCTACTTCACAGTGTCTTGGAGAGTATAATG 564 481 GTGATCCGTAATAATGGACAAATTCGAGTCTACTTCACAGTGTCTTGGAGAGTATAATG 540 565 AAACAATGCTTGAAGCTTTTCAGAGGAGCCCTCGGAAGCAGGCGCTGGCGCGCAG 624	Qy
ORIGIN	Query Match 79.2%; Score 1434.8; DB 6; Length 1614; Best Local Similarity 90.6%; Pred. No. 0; Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1; 1 AGTACACACGACCTGAGTGGAGTGGACACGAGAGCCCTTCCATGTTTAGGACCTCT 84 1 AGTACACACGACCTGAGTGGAGTGGACACGAGAGCCCTTCCATGTTTAGGACCTCT 60 85 GGGCTCAGAGAGTGGCGCGCGCCCTGGCGGACCTCCCGGATCGCGGGCGGAATG 144 61 GGGCTCAGAGAGTGGCGCGCGCCCTGGCGGACCTCCCGGATCGCGGGCGGAATG 120 145 GTCCGGTGGTCCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 204 121 GTCCGGTGGTCCCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180 205 GTTCGGTGGTTCGAGTTCGGGGCGCTTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 264 181 GTTCGGTGGTTCGAGTTCGGGGCGCTTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240 265 GTCCGTCCGGCGGGTCCCGGGAGCGGCTAGGCTCCCGGGTCCCGGGCGCCAGGC 324 241 GTCCGTCCGGCGGGTCCCGGGAGCGGCTAGGCTCCCGGGTCCCGGGCGCCAGGC 300 325 ATTCGGGGTGCAGATTGACGGGATCCCGGATGCAACCGCGCGCGCGCGCGCGCGCG 384 301 ATTCGGGGTGCAGATTGACGGGATCCCGGATGCAACCGCGCGCGCGCGCGCGCGCG 360 385 ACGGGTCCAGACCTGGTGGAGAGTGGGGAGCGGTCCCTGAGGATCCCGATGCGCT 444 361 ACGGGTCCAGACCTGGTGGAGAGTGGGGAGCGGTCCCTGAGGATCCCGATGCGCT 420 445 ACAGGCCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTACCTCAGTTTT 504 421 ACAGGCCAAGATGCTCAGCTTTATAGGTGTGACCTACACATGTGACCTTACCTCAGTTTT 480 505 GTGATCCGTAATAATGGACAAATTCGAGTCTACTTCACAGTGTCTTGGAGAGTATAATG 564 481 GTGATCCGTAATAATGGACAAATTCGAGTCTACTTCACAGTGTCTTGGAGAGTATAATG 540 565 AAACAATGCTTGAAGCTTTTCAGAGGAGCCCTCGGAAGCAGGCGCTGGCGCGCAG 624	Qy

QY 1705 TGTCTCAACCACTACTCAAAATTTAGTGGCAACCCAGGGGCTGGCACCTCCCAATCATCC 1764
 Db 1515 TGTCTCAACCACTACTCAAAATTTAGTGGCAACCCAGGGGCTGGCACCTCCCAATCATCC 1574

QY 1765 ATTCTCTGTCTGCGCAAGTGGCAATAAACGGGTGATTGCC 1804
 Db 1575 ATTCTCTGTCTGCGCAAGTGGCAATAAACGGGTGATTGCC 1614

RESULT 10
 AK027741
 LOCUS AK027741 1614 bp mRNA linear PRI 01-AUG-2002
 DEFINITION Homo sapiens cDNA FLJ4835 fis, clone OVARC1001369.
 ACCESSION AK027741
 VERSION AK027741.1 GI:14042642
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sato, H., Sugano, S., Shiratori, A., Sudo, H.,
 Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
 Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1614)
 Isogai, T. and Otsuki, T.
 Direct Submission
 Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; CDNA full insert sequencing;
 Research Association for Biotechnology; CDNA library construction;
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES
 source
 1..1614
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="OVARC1001369"
 /tissue_type="ovary, tumor tissue"
 /clone_lib="OVARC1"
 /note="cloning vector: pME18SF13"
 635..1168
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB55333.1"
 /db_xref="GI:14042643"
 /translation="MKTPVELAVSGMOTLGLHRCRGVYVKAPTSYVDETLEGPAG
 TRPTPDPPPPVVEKANRTRGVGKSKALGAGCETTPSRGCTPTLTTPRKQKRYR
 ISHTPSYCESLFGSRSEGSFAPRMAXGDAKRLALWTTPPTPRGSHSPRPREAP
 VPTQGLRISGLPRQG"
 1114..1115
 misc_feature
 /note="1116 bases segment is present in AY007158."

ORIGIN
 Query Match 79.2%; Score 1434.8; DB 9; Length 1614;
 Best Local Similarity 90.8%; Pred. No. 0;
 Matches 1612; Conservative 0; Mismatches 2; Indels 166; Gaps 1;
 QY 25 AGTACACAGCAGCCTGAGTGAAGTGGCACCAGAGGACCTCTCCATGTTTAGGACCTCCT 84
 Db 1 AGTACACAGCAGCCTGAGTGAAGTGGCACCAGAGGACCTCTCCATGTTTAGGACCTCCT 60

QY 85 GGGCCTCAGGACGCTGGCGCCGCCCTTGGCGGACTCCCGCCATCCGCGGCGCGAATG 144
 Db 61 GGGCCTCAGGACGCTGGCGCCGCCCTTGGCGGACTCCCGCCATCCGCGGCGCGAATG 120

QY 145 GTCCGGGTGCGGTCCGAGTGTCTCTGCTGGTGTCTCCCTGGTGTGGGTGCAAGTGTCTGG 204
 Db 121 GTCCGGGTGCGGTCCGAGTGTCTCTGCTGGTGTCTCCCTGGTGTGGGTGCAAGTGTCTGG 180

QY 205 GTTCTGGGTCTTCTGATTCGCGGGCGCTTACAGTAGCTGTGCGGGTCTCTCGGGTGA 264
 Db 181 GTTCTGGGTCTTCTGATTCGCGGGCGCTTACAGTAGCTGTGCGGGTCTCTCGGGTGA 240

QY 265 GTCCGTCCGCGCGGGTGTCCCGCGGAGCGCTTAGGCTGCGCGGGGTCCGCGGGGCCCAAGGC 324
 Db 241 GTCCGTCCGCGCGGGTGTCCCGCGGAGCGCTTAGGCTGCGCGGGGTCCGCGGGGCCCAAGGC 300

QY 325 ATTCCGGGTCCAGATTGACGGGGATCCGGATGACCGGGTCCCGGGTCCCGGGTCCCGG 384
 Db 301 ATTCCGGGTCCAGATTGACGGGGATCCGGATGACCGGGTCCCGGGTCCCGGGTCCCGG 360

QY 385 ACGGTCCAGACCTGGTGGGAGAGGTGGGGAGCGGTCCCTGAGGATCCCGATCCCT 444
 Db 361 ACGGTCCAGACCTGGTGGGAGAGGTGGGGAGCGGTCCCTGAGGATCCCGATCCCT 420

QY 445 ACGAGCCAAAGTGTCTAGCTTTATAGTGTGACCTACACATGTGACTTACCTCAGTTT 504
 Db 421 ACGAGCCAAAGTGTCTAGCTTTATAGTGTGACCTACACATGTGACTTACCTCAGTTT 480

QY 505 GTGATCCGTAAATGGACAAATTCGAAGTTCGAGTGTCTTACAGTGTGTTGAGAGGATTAATG 564
 Db 481 GTGATCCGTAAATGGACAAATTCGAAGTTCGAGTGTGTTGAGAGGATTAATG 540

QY 565 AAAAATGCTTGTAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTCGCCCGGAG 624
 Db 541 AAAAATGCTTGTAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTCGCCCGGAG 600

QY 625 AGCACCTCTGTACACAGGAGCCACAGCAGCAGTGAAGACCCCGTGGAGTGGCCGT 684
 Db 601 AGCACCTCTGTACACAGGAGCCACAGCAGCAGTGAAGACCCCGTGGAGTGGCCGT 660

QY 685 CAGTGGGATGCAGACCTCCGCTTTCAGACCCCTTCCGAGGTGGCTACCGGTCAAGGC 744
 Db 661 CAGTGGGATGCAGACCTCCGCTTTCAGACCCCTTCCGAGGTGGCTACCGGTCAAGGC 720

QY 745 CAGGACCTCATATGTGGATGAGACTCTGTTTGGCAGCCAGAGGACCCCGGCTTACCC 804
 Db 721 CAGGACCTCATATGTGGATGAGACTCTGTTTGGCAGCCAGAGGACCCCGGCTTACCC 780

QY 805 ACCGACTTCGATCCGCTTGGGTGGAGAGGCTTAACAGAACCCAGAGGCTGGGCAAGGA 864
 Db 781 ACCGACTTCGATCCGCTTGGGTGGAGAGGCTTAACAGAACCCAGAGGCTGGGCAAGGA 840

QY 865 GGCATCAAGGCTTGGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGC 924
 Db 841 GGCATCAAGGCTTGGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGC 900

QY 925 CCCACCCCTCACACCCAGGAGGAGAACAAATACAGCCCATCAGCACACCCCGCTTA 984
 Db 901 CCCACCCCTCACACCCAGGAGGAGAACAAATACAGCCCATCAGCACACCCCGCTTA 960

QY 985 CTGTGATGAGTCTGTTTGGCTCCCGATCTGAAGGCGCCAGCTTCCGGGCGCCCGGAT 1044
 Db 961 CTGTGATGAGTCTGTTTGGCTCCCGATCTGAAGGCGCCAGCTTCCGGGCGCCCGGAT 1020

QY 1045 GCGGAGGGGATGCGCCAGAGCTCGTGTCTCTTGTGAGCCACCACTACCCCG 1104
 Db 1021 GCGGAGGGGATGCGCCAGAGCTCGTGTCTCTTGTGAGCCACCACTACCCCG 1080

QY 1105 GGGTAGGCATCGCCCGCGCCCGCAGGAGGACCACTGCGAGGCATTCACCCAGCTGCTCC 1164
 Db 1081 GGGTAGGCATCGCCCGCGCCCGCAGGAGGAGG-----1109

Db 121 GATTAAATGAAACAAATGCTTGTAAAGCTCTTTGACAGGAGGAGCTTCGGAGCAGGGCCT 180
Qy 616 GCGCGGAGAGACACACCTGCTGTGTACACAGGAGACACAGGAGCATGAAAGACCCCGGTGGA 675
Db 181 GCGCGGAGAGACACACCTGCTGTGTACACAGGAGACACAGGAGCATGAAAGACCCCGGTGGA 240
Qy 676 GCTGGCGGTAGTGGGATGAGACACCTCGGCGCTTTCAGCACCCTGCGGAGGTGCTACCG 735
Db 241 CTTGGCGGTAGTGGGATGAGACACCTCGGCGCTTTCAGCACCCTGCGGAGGTGCTACCG 300
Qy 736 GGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCCGCAGGACACCG 795
Db 301 GGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCCGCAGGACACCG 360
Qy 796 GCCTACCCCGGACGTCGATCCGCGCTTGGTGGAGAGGCTTAAAGACACGAGCGT 855
Db 361 GCCTACCCCGGACGTCGATCCGCGCTTGGTGGAGAGGCTTAAAGACACGAGCGT 420
Qy 856 GGGCAAGGAGGATCGAAGGCTTGGGGGCAAGGGGAGCTGTGAGACACACCCCTCAAG 915
Db 421 GGGCAAGGAGGATCGAAGGCTTGGGGGCAAGGGGAGCTGTGAGACACACCCCTCAAG 480
Qy 916 GGGCAGACACCCGACCTTCACACGAGGAAGAGAAACAAATACAGACCCCATAGCACAC 975
Db 481 GGGCAGACACCCGACCTTCACACGAGGAAGAGAAACAAATACAGACCCCATAGCACAC 540
Qy 976 CCCGCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAGGGCGGAGCTTCGGGGC 1035
Db 541 CCCGCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAGGGCGGAGCTTCGGGGC 600
Qy 1036 CCCGCGATGGGAGGGGATGCGCAAGCTCGGCTCTTTGTGAGCGCCACAC 1095
Db 601 CCCGCGATGGGAGGGGATGCGCAAGCTCGGCTCTTTGTGAGCGCCACAC 660
Qy 1096 TACCCCGAGGGTACGACCTCGCCCGCCCGAGGAGGACCACTCGGAGCAATCACCC 1155
Db 661 TACCCCGAGGGTACGACCTCGCCCGCCCGAGGAGGACCACTCGGAGCAATCACCC 720
Qy 1156 AGCTGGTCCCTTCCAGACAGAGCGGGGCGAGGAGCTCCAGAGTTATCTATGG 1215
Db 721 AGCTGGTCCCTTCCAGACAGAGCGGGGCGAGGAGCTCCAGAGTTATCTATGG 780
Qy 1216 TGGGTTACATCTTCAGCGCCCTGAGCGGGACCTTCCTTCCTCACCCACCTGAA 1275
Db 781 TGGGTTACATCTTCAGCGCCCTGAGCGGGACCTTCCTTCCTCACCCACCTGAA 840
Qy 1276 TGTCCCGAGCACTGGTTCATCCAGCCACGAGTGGCCCGCCACACAAATGGGCTCAGGATCT 1335
Db 841 TGTCCCGAGCACTGGTTCATCCAGCCACGAGTGGCCCGCCACACAAATGGGCTCAGGATCT 900
Qy 1336 CAGCGCTTCAGCTCAGGGTACCTTCGAGAGCCCGCTTGTGACTTCAGGGCTCGCTC 1395
Db 901 CAGCGCTTCAGCTCAGGGTACCTTCGAGAGCCCGCTTGTGACTTCAGGGCTCGCTC 960
Qy 1396 AGTTAGCATTTAGTGCATCTACCCAGCAGGAGTGGGCGCCACCAAGAACCAAGCC 1455
Db 961 AGTTAGCATTTAGTGCATCTACCCAGCAGGAGTGGGCGCCACCAAGAACCAAGCC 1020
Qy 1456 CCCTTGAATATGACTCTTTCATCAGGGTGTGCTATGGGGCCACCGGACAGGTATGGC 1515
Db 1021 CCCTTGAATATGACTCTTTCATCAGGGTGTGCTATGGGGCCACCGGACAGGTATGGC 1080
Qy 1516 CCCTTGCAGGTAGGAGGACATTCATCACCAGGGAACCCAGGTATTAAGAGAGCC 1575
Db 1081 CCCTTGCAGGTAGGAGGACATTCATCACCAGGGAACCCAGGTATTAAGAGAGCC 1140
Qy 1576 TGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCTTTATCTGCAATCTCT 1635
Db 1141 TGTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCTTTATCTGCAATCTCT 1200
Qy 1636 AGTCGATCTTGTCTTTCTCCGATTCGGGATTTGGGGGCGACCTCTAAGATGCTCT 1695
Db 1201 AGTCGATCTTGTCTTTCTCCGATTCGGGATTTGGGGGCGACCTCTAAGATGCTCT 1260

Qy 1696 CTCGAGCCCTGTCTCAACCATCTCCAAATAGTGTGCAACCCAGGGGCTGGCACTCCC 1755
Db 1261 CTCGAGCCCTGTCTCAACCATCTCCAAATAGTGTGCAACCCAGGGGCTGGCACTCCC 1320
Qy 1756 ACATCATCATCTGCTTCTGCTCCAGTCCGAATAAAGCGGCTGATTGCGCA 1806
Db 1321 ACATCATCATCTGCTTCTGCTCCAGTCCGAATAAAGCGGCTGATTGCGCA 1371

RESULT 12
AC089999/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Db 59708 CAGACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGTGTGGTCCCGATCTGA 59649
 QY 1018 AGGCGCCAGCTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCAAGACTCCGTGCTCT 1077
 Db 59648 AGGCGCCAGCTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCAAGACTCCGTGCTCT 59589
 QY 1078 CTTGTGGACGCCACCACTACCCACAGGGGTAGCCACTTCGCGCCGCCCGCCAGGAGGACCC 1137
 Db 59588 CTTGTGGACGCCACCACTACCCACAGGGGTAGCCACTTCGCGCCGCCCGCCAGGAGGACCC 59529
 QY 1138 ACTGGAGCCATTACCCAGTGTGCTCTCAAGACAGAGCGGGGACAGCGGAGACTC 1197
 Db 59528 ACTGGAGCCATTACCCAGTGTGCTCTCAAGACAGAGCGGGGACAGCGGAGACTC 59469
 QY 1198 CCAGAAGTTATCTATGGTGGGTACACTCTTCACGCGCCCTCGAAGCGGGGACTTTCCCA 1257
 Db 59468 CCAGAAGTTATCTATGGTGGGTACACTCTTCACGCGCCCTCGAAGCGGGGACTTTCCCA 59409
 QY 1258 TTCCTTACCCACCTGAATGTCCTCAGACCTGTGTATCCAGGACCACTAGTCCGCCCCACAC 1317
 Db 59408 TTCCTTACCCACCTGAATGTCCTCAGACCTGTGTATCCAGGACCACTAGTCCGCCCCACAC 59349
 QY 1318 AAATGGGCTCAGGATCTCAGGCGCTTCCAGTCAAGGGGTGACCTTCCGGAGCCCTTGGT 1377
 Db 59348 AAATGGGCTCAGGATCTCAGGCGCTTCCAGTCAAGGGGTGACCTTCCGGAGCCCTTGGT 59289
 QY 1378 GACTTCCAGGCTCGCTCAGTGTAGCAATTCAGTGCCATCTACCCACAGAGGTGGGC 1437
 Db 59288 GACTTCCAGGCTCGCTCAGTGTAGCAATTCAGTGCCATCTACCCACAGAGGTGGGC 59229
 QY 1438 CACCCAGAAACCAAGCCCTTGGCAATGATATCTCTTTCATCAGGGTTCCTATGGGC 1497
 Db 59228 CACCCAGAAACCAAGCCCTTGGCAATGATATCTCTTTCATCAGGGTTCCTATGGGC 59169
 QY 1498 CACGGCAGCAGGTATGCGCCCTTGGCAAGGTAGGAGACATTCATCACCCAGGAAACCC 1557
 Db 59168 CACGGCAGCAGGTATGCGCCCTTGGCAAGGTAGGAGACATTCATCACCCAGGAAACCC 59109
 QY 1558 AGGTATTAAAGAGCCCTGTGGGGGACAGACATAGCAGGGGTGGGAGTCCCTCC 1617
 Db 59108 AGGTATTAAAGAGCCCTGTGGGGGACAGACATAGCAGGGGTGGGAGTCCCTCC 59049
 QY 1618 TTTATCTGACATCTCTAGTGCATCTTCCCTTTTCTCCGATTCGATTTGGGGC 1677
 Db 59048 TTTATCTGACATCTCTAGTGCATCTTCCCTTTTCTCCGATTTGGGGC 58989
 QY 1678 CACTCTAAGATGCTCTCCAGCCCTGTCTCAACATCTCAAAATAGTGCCAAACC 1737
 Db 58988 CACTCTAAGATGCTCTCTCCAGCCCTGTCTCAACATCTCAAAATAGTGCCAAACC 58929
 QY 1738 AGGGGCTGACCTCCACATCATCTCTCTGCTCCAGTCCGATTAACGGGT 1797
 Db 58928 AGGGGCTGACCTCCACATCATCTCTCTGCTCCAGTCCGATTAACGGGT 58869
 QY 1798 GATTGCCAACCTGG 1811
 Db 58868 GATTGCCAACCTGG 58855

RESULT 13
 AX869030
 LOCUS
 DEFINITION Sequence 3935 from Patent EP1074617.
 ACCESSION AX869030
 VERSION AX869030.1 GI:40023893
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

TITLE Primers for synthesizing full-length cDNA and their use
 JOURNAL Patent: EP 1074617-A 3935 07-FEB-2001;
 Research Association for Biotechnology (JP)
 FEATURES
 source
 1. .784
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 36.5%; Score 661.4; DB 6; Length 784;
 Best Local Similarity 98.7%; Pred. No. 2.6e-139;
 Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
 QY 191 GTGCAAGTGTGGGTTCTGGGTTCTGGATTTCGGGGCCGTTTCACAGTAGCCTGTGCC 250
 Db 1 GTGCAAGTGTGGGTTCTGGGTTCTGGATTTCGGGGCCGTTTCACAGTAGCCTGTGCC 60
 QY 251 GGTCTCTCGGTGATGTCGTCCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGT 310
 Db 61 GGTCTCTCGGTGATGTCGTCCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGT 120
 QY 311 CCGGGGCCCGAGGCATTCGCGGCTGCGAGATTGACGGGATCCCGGATGACCGCGCGGCC 370
 Db 121 CCGGGGCCCGAGGCATTCGCGGCTGCGAGATTGACGGGATCCCGGATGACCGCGCGGCC 180
 QY 371 CCGGGCCCTCACCCAGGGTCCAGACTGTGGGGAAGGTGCGGGGACGGGTCCCTGA 430
 Db 181 CCGGGCCCTCACCCAGGGTCCAGACTGTGGGGAAGGTGCGGGGACGGGTCCCTGA 240
 QY 431 GGATCCGATGCTTACGAGCCCAAGATGCTCAGCTTTATAGTGTGACCTACATGTGAC 490
 Db 241 GGATCCGATGCTTACGAGCCCAAGATGCTCAGCTTTATAGTGTGACCTACATGTGAC 300
 QY 491 TTCACTCAGTTTGTGATCGTAAATGGACAAATTCGAAGCTACTTTCAGTGTCTGTT 550
 Db 301 TTCACTCAGTTTGTGATCGTAAATGGACAAATTCGAAGCTACTTTCAGTGTCTGTT 360
 QY 551 GAGAGGATTAATGAACAAATGCTTTGAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAG 610
 Db 361 GAGAGGATTAATGAACAAATGCTTTGAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAG 420
 QY 611 GGCCTGCGCGCAGACACACCTGCTGTACCAGGAGACACAGGACATGAAGACCCC 670
 Db 421 GGCCTGCGCGCAGACACACCTGCTGTACCAGGAGACACAGGACATGAAGACCCC 480
 QY 671 GTGAGCTGCGCGTCACTGGGATGACAGACCTTGGGCTTCAGCACCCCTGCGAGGTGGC 730
 Db 481 GTGAGCTGCGCGTCACTGGGATGACAGACCTTGGGCTTCAGCACCCCTGCGAGGTGGC 540
 QY 731 TACCGGTCAAGGCAGAGCTGATATGTGGATG-AGACTCTGTTGGAGCCAGCAGG 789
 Db 541 TACCGGTCAAGGCAGAGCTGATATGTGGATGAGACTCTGTTGGAGCCAGCAGG 600
 QY 790 CACCGGCTACCCACCGACTTCGATCCGCTGCGGTGAGAGGCTAACAGAACCCAG 849
 Db 601 CACCGGCTACCCACCGACTTCGATCCGCTGCGGTGAGAGGCTAACAGAACCCAG 660
 QY 850 AGCGTGGGCAAGGA-GGCTTCGAAGGCTTTGGGG 884
 Db 661 AGCGTGGGCAAGGAGGCTTCGAAGGCTTTGGGG 696

RESULT 14
 BD149092
 LOCUS
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD149092
 VERSION BD149092.1 GI:27854850
 KEYWORDS JP 2002191363-A/3935.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 784)
Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 3935 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/3935
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SATOH,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CO
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..784
/organism='Homo sapiens (human)'
/feature=
Location/Qualifiers
1..784
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Query Match 36.5%; Score 661.4; DB 6; Length 784;
Best Local Similarity 98.7%; Pred. No.2.6e-139;
Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 191 GTGCAAGTCTGGGTTCGGTTCTCGATTTCGGATTCGGGGCGGTCACACGTAGCCTGTGCC 250
Db 1 GTGCAAAAGTGCTGGGTTCGGATTTCGGATTCGGGGCCGTTTACACGTAGCCTGTGCC 60

Qy 251 GGCTCTCTCGGTGAGTGCCTCCGGCGCGGGTCTCGGACGGCTTAGCTGCCGGGGGT 310
Db 61 GGCTCTCTCGGTGAGTGCCTCCGGCGCGGGTCTCGGACGGCTTAGCTGCCGGGGGT 120

Qy 311 CCGGGGCCCGCAGGACTTCGGGGCTGCAGATTGACGGGGATCCCAGGTGACCGCGCCCC 370
Db 121 CCGGGGCCCGCAGGACTTCGGGGCTGCAGATTGACGGGGATCCCAGGTGACCGCGCCCC 180

Qy 371 CCGGGCCCTCACCGACGGGTCCAGACTGTGTGGGAAGAAGTGTGGGGACGGGTCCCTGA 430
Db 181 CCGGGCCCTCACCGACGGGTCCAGACTGTGTGGGAAGAAGTGTGGGGACGGGTCCCTGA 240

Qy 431 GGATCCCGATCCCTACGAGCCAAGATGCTCAGCTTTATAGGTGTGACCTACATGTGAC 490
Db 241 GGATCCCGATCCCTACGAGCCAAGATGCTCAGCTTTATAGGTGTGACCTACATGTGAC 300

Qy 491 TTCACCTCAGTTTTGTGATCCGTAAAATGGACAATAATCGAAGCTACTTTCAGAGTCTGTT 550
Db 301 TTCACCTCAGTTTTGTGATCCGTAAAATGGACAATAATCGAAGCTACTTTCAGAGTCTGTT 360

Qy 551 GAGAGGATTAATGAAACAATGCTTTGTAAGCTTTTTCAGAGAGGAGCCTCGGAAGCAG 610
Db 361 GAGAGGATTAATGAAACAATGCTTTGTAAGCTTTTTCAGAGAGGAGCCTCGGAAGCAG 420

Qy 611 GGCCTGGCCGCGAGAGCAACCTGCTGTTCACCAGGACCAACAGGCAGCATGAGACCCCC 670
Db 421 GGCCTGGCCGCGAGAGCAACCTGCTGTTCACCAGGACCAACAGGCAGCATGAGACCCCC 480

Qy 671 GTGAGCTGTGGCGGTTCAGTGGGATGAGACCCCTCGGCCCTTCAGCACCGCTGCCAGAGTGCC 730
Db 481 GTGAGCTGTGGCGGTTCAGTGGGATGAGACCCCTCGGCCCTTCAGCACCGCTGCCAGAGTGCC 540

Qy 731 TACCGGGTCAAGGCCAGGACGTTCATATGTGGATG-AGACTCTGTTTGGCAGCCAGCAGG 789
Db 541 TACCGGGTCAAGGCCAGGACGTTCATATGTGGATGAGACTCTGTTTGGCAGCCAGCAGG 600

```

[illegible]

QY 564 GAAACAATGCTTGTAAAGCTCTTTGCAGGAGGAGCCTTCGGAAGCAGGGCCTGGCC 619
|||
Db 541 GAAACAATGCTGTAAAGCTCTTTGCAGGAGGAGCCTTCGGAAGCAGGGCCTGGCC 596
|||

Search completed: August 17, 2004, 08:21:00
Job time : 7227 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 03:21:14 ; Search time 873 Seconds
(without alignments)
8812.700 Million cell updates/sec

Title: US-10-031-589-3
Perfect score: 1811
Sequence: 1 agtgcctggcgcctcgcg.....cggcgtgattgccaactgg 1811

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 'N Geneseq_29Jan04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1811	100.0	1811	4 AAF58163	Aaf58163 Human neu
2	1621	89.5	1775	4 AAHL6354	Aahl6354 Human CDN
3	1610.2	88.9	1826	4 ABA08585	Aba08585 Human typ
4	1434.8	79.2	1614	4 AAHL6384	Aahl6384 Human CDN
5	661.4	36.5	784	4 AAH07100	Aah07100 Human CDN
6	564.4	31.2	889	5 AAS75219	Aas75219 DNA encod
7	562	31.0	597	4 AAH07172	Aah07172 Human CDN
8	494	27.3	561	4 AAHL1267	Aahl1267 Human CDN
9	436.6	24.1	477	8 ACH14455	Achl14455 Human adu
10	410.6	22.7	484	8 ACH43954	Ach43954 Human foe
11	341.8	18.9	578	4 AAHL1203	Aahl1203 Human CDN
12	330	18.2	440	5 AAS75218	Aas75218 DNA encod
13	138	7.6	494	9 ADB51383	Adb51383 Primari r
14	130.8	7.2	292	6 ABR79266	Abr79266 Human ORF
15	60	3.3	60	6 ABR32144	Abr32144 Human spl
16	54	3.0	2000	7 ADA71938	Ada71938 Rice gene
17	51.6	2.8	680	4 AA199338	Aa199338 Human exc
18	51.6	2.8	680	4 AA199337	Aa199337 Human exc
19	51.6	2.8	680	5 AA163687	Aa163687 Human kid
20	51.6	2.8	680	5 AA163688	Aa163688 Human kid
21	50.8	2.8	343	4 AA198796	Aa198796 Human exc
22	50.8	2.8	343	5 AA163192	Aa163192 Human kid
23	50.4	2.8	43981	9 ADC86790	Adc86790 Human GPC

24	50	2.8	316	6 ABL84296	Ab184296 Human ova
25	49.8	2.7	174493	7 ACA61395	Aca61395 Novel hum
26	49.8	2.7	174493	9 AAD59937	Aad59937 Human kin
27	48.4	2.7	267	6 ABL84293	Ab184293 Human ova
28	48.4	2.7	304	6 ABL84286	Ab184286 Human ova
29	47	2.6	110000	7 ABZ79565_3	Continuation (4 of
30	46.6	2.6	342	4 AAL2164	Aal12164 Human bre
31	46.6	2.6	566	4 AAL21047	Aal21047 Human bre
32	46.6	2.6	1964	4 AAF32803	Aaf32803 Human sec
33	46.2	2.6	12111	5 ABA21422	Aba21422 Human ner
34	46.2	2.6	30001	2 AAT61016	Aat61016 Total DNA
35	45.2	2.6	30001	2 AAX05110	Aax05110 S. aureof
36	45.8	2.5	172637	6 ABR83124	Abn83124 Human vol
37	45.8	2.5	349901	9 ADC86940	Adc86940 Human GPC
38	45.6	2.5	2000	7 ADA71938	Ada71938 Rice gene
39	45.6	2.5	72928	2 AAZ18355	Aaz18355 Human AST
40	45.6	2.5	72928	3 AAA80253	Aaa80253 Human AST
41	45.6	2.5	114955	2 AAX53491	Aax53491 Human ade
42	45.4	2.5	22756	4 AAL03921	Aal03921 Human rep
43	45.4	2.5	22756	5 AAS40321	Aas40321 DNA encod
44	45.2	2.5	10732	3 AAA10594	Aaa10594 Gene enco
45	45	2.5	13069	4 AAS36350	Aas36350 Human car

ALIGNMENTS

RESULT 1
AAF58163
ID AAF58163 standard; DNA; 1811 BP.
AC AAF58163;
XX
DT 23-APR-2001 (first entry)
DE Human neuron progenitor cell clone #2 DNA.
XX
KW Neuron; progenitor cell; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WC200107607-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-JP004895.
XX
PR 23-JUL-1999; 99JP-00209817.
PR 18-OCT-1999; 99US-0159528P.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
DR WPI; 2001-182791/18.
XX
PT New human polynucleotides, particularly DNAs, isolated from a cDNA
PT library derived from progenitor cells, useful in gene therapy, as well as
PT in producing proteins useful as diagnostic markers in drug development.
XX
PS Claim 1; Page 33-36; 54pp; English.
XX
CC The present invention relates to human proteins isolated from clones from
CC neuron progenitor cells. The proteins and the DNA encoding them may be
CC used in the preparation of treatments for diseases associated with the
CC proteins
XX
SQ Sequence 1811 BP; 355 A; 588 C; 535 G; 333 T; 0 U; 0 Other;

Query Match 100.0%; Score 1811; DB 4; Length 1811;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTGCCTCGGGCCCTCGGGCCCTAGTACACACGCACTGAGTGAGTGGCAACAGAGGAC 60
Db 1 AGTGCCTCGGGCCCTCGGGCCCTAGTACACACGCACTGAGTGAGTGGCAACAGAGGAC 60
QY 61 CCTCTCCATGTTTAGGACCTCTCTGGGCTCAGAGAGGTGGCGCCCGCCCTCGGGCGGAC 120
Db 61 CCTCTCCATGTTTAGGACCTCTCTGGGCTCAGAGAGGTGGCGCCCGCCCTCGGGCGGAC 120
QY 121 TCCGCCCATCCGGGCGCGCAATGCTCGGGTCCGGTCCGAGTGTCTGCTGCTGCTCC 180
Db 121 TCCGCCCATCCGGGCGCGCAATGCTCGGGTCCGGTCCGAGTGTCTGCTGCTGCTCC 180
QY 181 TGGTGTCTGGGTGCAAAAGTGTCTGGGTCTGGGTCTGGATTCGCGGGCCGTTCCACAGT 240
Db 181 TGGTGTCTGGGTGCAAAAGTGTCTGGGTCTGGGTCTGGATTCGCGGGCCGTTCCACAGT 240
QY 241 AGCCTGTGCCGGCTCCTCGGTGAGTCCGTCCGGCGCGGTGCCCGGGACGGCCTAGGC 300
Db 241 AGCCTGTGCCGGCTCCTCGGTGAGTCCGTCCGGCGCGGTGCCCGGGACGGCCTAGGC 300
QY 301 TCCGGGGGTCCGGGCCCCAGGCAATCCGGGCTGCAGATGACGGGATCCCGGATGCA 360
Db 301 TCCGGGGGTCCGGGCCCCAGGCAATCCGGGCTGCAGATGACGGGATCCCGGATGCA 360
QY 361 CCGCGCGCCCCCGGCCCTCACCCAGCGGTCCAGACCTGGTGGGAAGAAGTCCGGGGAC 420
Db 361 CCGCGCGCCCCCGGCCCTCACCCAGCGGTCCAGACCTGGTGGGAAGAAGTCCGGGGAC 420
QY 421 GGGTCCCTGAGGATCCCGATGCCCTACAGCCCAAGATGCTCAGCTTTATAGGTGTACCTA 480
Db 421 GGGTCCCTGAGGATCCCGATGCCCTACAGCCCAAGATGCTCAGCTTTATAGGTGTACCTA 480
QY 481 CACATGTGACTTCACTCAGTTTGTGATCCGTAAATGACAAATTCGAAGCTACTTCA 540
Db 481 CACATGTGACTTCACTCAGTTTGTGATCCGTAAATGACAAATTCGAAGCTACTTCA 540
QY 541 CAGTGTGTTGAGAGGATTAATGAACAAATGTTGTAAGTCTTTGAGGAGGAGCC 600
Db 541 CAGTGTGTTGAGAGGATTAATGAACAAATGTTGTAAGTCTTTGAGGAGGAGCC 600
QY 601 TCGAAGCAGGGCTGCGCGGACAGCACACCTGCTCAGCAGGACACACAGGAGCAT 660
Db 601 TCGAAGCAGGGCTGCGCGGACAGCACACCTGCTCAGCAGGACACACAGGAGCAT 660
QY 661 GAAGACCCCGGTGAGAGTGGCCCTCAGTGGGATGACAGACCTCGGCCCTTTCAGCACCGCTG 720
Db 661 GAAGACCCCGGTGAGAGTGGCCCTCAGTGGGATGACAGACCTCGGCCCTTTCAGCACCGCTG 720
QY 721 CCGAGTGGGTACCGGGTCAAGCCAGGAGCTCATATGTGGATGAGACTCTGTTGGCAG 780
Db 721 CCGAGTGGGTACCGGGTCAAGCCAGGAGCTCATATGTGGATGAGACTCTGTTGGCAG 780
QY 781 CCGAGGAGGACACCGGCTTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGCTAA 840
Db 781 CCGAGGAGGACACCGGCTTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGCTAA 840
QY 841 CAGAACCCAGAGGGTGGGCAAGGAGGATCGAAGGCTTTGGGGCAAGAGGAGCTGTGA 900
Db 841 CAGAACCCAGAGGGTGGGCAAGGAGGATCGAAGGCTTTGGGGCAAGAGGAGCTGTGA 900
QY 901 GACACCCCTCAAGGGGAGCACCCACCTCACACCAAGGAGGAGCAACAAATACAG 960
Db 901 GACACCCCTCAAGGGGAGCACCCACCTCACACCAAGGAGGAGCAACAAATACAG 960
QY 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGTCTCCCGATCTGAAGG 1020
Db 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGTCTCCCGATCTGAAGG 1020
QY 1021 CGCAGCTTCGGGGCCCCCGGATGCGAAGGGGATGCGCAAGCTCCGTCCTCTT 1080
Db 1021 CGCAGCTTCGGGGCCCCCGGATGCGAAGGGGATGCGCAAGCTCCGTCCTCTT 1080
QY 1081 GTGGACGCCACCACTTACCCCGAGGGGTAGCCACTCGCCCGCCCGGAGGAGGCCACCT 1140

Db 1081 GTGGACGCCACCACTTACCCCGAGGGGTAGCCACTCGCCCGCCCGGAGGAGGCCACCT 1140
QY 1141 GGGAGCCATTACCCAGCTGGTCCCTCAAGACAGAGCCGGGGCCAGCGGAGACTCCCA 1200
Db 1141 GGGAGCCATTACCCAGCTGGTCCCTCAAGACAGAGCCGGGGCCAGCGGAGACTCCCA 1200
QY 1201 GAAGTTATCTATGGGTGGGTACACTCTTCAAGCCCCCTGAAGCGGGACTTTCCCATTC 1260
Db 1201 GAAGTTATCTATGGGTGGGTACACTCTTCAAGCCCCCTGAAGCGGGACTTTCCCATTC 1260
QY 1261 CCTCACCCACCTGATGTCCTCCAGCAGCTGTCATCCAGCACCAAGTGCCTCCACAAA 1320
Db 1261 CCTCACCCACCTGATGTCCTCCAGCAGCTGTCATCCAGCACCAAGTGCCTCCACAAA 1320
QY 1321 TGGGCTTCAGGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGGAGCCCTCTGGTAC 1380
Db 1321 TGGGCTTCAGGATCTCAGGCTTCCACGTCAGGGGTGACCTTCCGGAGCCCTCTGGTAC 1380
QY 1381 TTCAGGGCTCCTCAGTTAGCATTTTCAAGTCCCATCTACCCACGACGAGTGGGGCCAC 1440
Db 1381 TTCAGGGCTCCTCAGTTAGCATTTTCAAGTCCCATCTACCCACGACGAGTGGGGCCAC 1440
QY 1441 CCAGAAACAAAGCCCCCTTGGAAATGATCTTTTCAATCAGGGTTGCCCTATGGGGCCAC 1500
Db 1441 CCAGAAACAAAGCCCCCTTGGAAATGATCTTTTCAATCAGGGTTGCCCTATGGGGCCAC 1500
QY 1501 GGGCAGAGTATGGCCCTTGGCAGGGTAGAGGACATTCATCACCCAGGGAACCCGAG 1560
Db 1501 GGGCAGAGTATGGCCCTTGGCAGGGTAGAGGACATTCATCACCCAGGGAACCCGAG 1560
QY 1561 TATTAAAGAACCCCTGTTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTT 1620
Db 1561 TATTAAAGAACCCCTGTTGGGGCAGACAGACATAGCAGGGGTGGGAGTGCCTCCCTTT 1620
QY 1621 ATCTGACAATCTCTAGTCGATTTTCTCCCTTTTCTCCCGATTTGGGGGCGCAC 1680
Db 1621 ATCTGACAATCTCTAGTCGATTTTCTCCCTTTTCTCCCGATTTGGGGGCGCAC 1680
QY 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCAATCTCATAATAGTGCACCCAG 1740
Db 1681 CTCTAAGATGCTCTCTCCAGCCCTGTCTCAACCAATCTCATAATAGTGCACCCAG 1740
QY 1741 GGCCTGGCAGCTCCACATCATCTTGTCTTGTGCAAGTGCATAAAGCGGCGTAT 1800
Db 1741 GGCCTGGCAGCTCCACATCATCTTGTCTTGTGCAAGTGCATAAAGCGGCGTAT 1800
QY 1801 TGGCAACCTGG 1811
Db 1801 TGGCAACCTGG 1811

RESULT 2

AAH16354

ID AAH16354 standard; cDNA; 1775 BP.

XX

AC AAH16354;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA sequence SEQ ID NO:15283.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

FN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.


```
QY 145 GTCCGGGTCCGGTCCGAGTGTCTGCTGGCTGCTCCCTGGTGTGCTGGTGCAAAAGTCTGG 204
DB 121 GTCCGGGCGGGTCCGAGTGTCTGCTGGCTGCTCCCTGGTGTGCTGGTGCAAAAGTCTGG 180
QY 205 GTTCTGGGTTCTCGATTCCGGGCGGTTTACAGTAGCTGTGCGGCTCTCTCGAGTGA 264
DB 181 GTTCTGGGTTCTCGATTCCGGGCGGTTTACAGTAGCTGTGCGGCTCTCTCGAGTGA 240
QY 265 GTCCGTCCGGCGCGGTGCTCCCGGACGGCTTAGCTGTCCGGGGTCCGGGGCCCCAGGC 324
DB 241 GTCCGTCCGGCGCGGTGCTCCCGGACGGCTTAGCTGTCCGGGGTCCGGGGCCCCAGGC 300
QY 325 ATTCCGGGCTGCAGATTGACGGGGATCCGGGATCCGGATCCAGCGCGCGCCCGCTCACCG 384
DB 301 ATTCCGGGCTGCAGATTGACGGGGATCCGGGATCCAGCGCGCGCCCGCTCACCG 360
QY 385 ACCGGTCCAGACCTGGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCCGATGCCCT 444
DB 361 ACCGGTCCAGACCTGGTGGGAAGAGTGGGGGACGGGTCCCTGAGGATCCCGATGCCCT 420
QY 445 ACAGACCAAGATGCTAGCTTTATAGGTGTGACTTACATGATGACTTCACTCACTT 504
DB 421 ACAGACCAAGATGCTAGCTTTATAGGTGTGACTTACATGATGACTTCACTCACTT 480
QY 505 GTGATCCGTAAATGGACAATTCGAAGTCTTACAGTGTGTTGAGAGGATTAATG 564
DB 481 GTGATCCGTAAATGGACAATTCGAAGTCTTACAGTGTGTTGAGAGGATTAATG 540
QY 565 AAACAATGCTTGTAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTTGGCCGGCAG 624
DB 541 AAACAATGCTTGTAAAGCTCTTTGAGGAGGAGCTTCGGAAGCAGGGCTTGGCCGGCAG 600
QY 625 AGCACACTGCTGTACAGGACACACAGCAGATGAGACCCCGTGGAGCTGGCCGT 684
DB 601 AGCACACTGCTGTACAGGACACACAGCAGATGAGACCCCGTGGAGCTGGCCGT 660
QY 685 CAGTGGGATGAGACCTCTGGCCCTTCAGACCCGCTGCGGAGTGGGTACCGGCTCAAGC 744
DB 661 CAGTGGGATGAGACCTCTGGCCCTTCAGACCCGCTGCGGAGTGGGTACCGGCTCAAGC 720
QY 745 CAGGACGTATATGTGATGAGACTCTGTTTGGAGCCAGCAGGACCCGGCTACCCC 804
DB 721 CAGGACGTATATGTGATGAGACTCTGTTTGGAGCCAGCAGGACCCGGCTACCCC 780
QY 805 ACCGACTTTCGATCCGCTGGGTGGAGAGGCTTAACAGAAACAGAGGCGTGGGCAAGCA 864
DB 781 ACCGACTTTCGATCCGCTGGGTGGAGAGGCTTAACAGAAACAGAGGCGTGGGCAAGCA 840
QY 865 GGCAATGAAGGCTTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGCAC 924
DB 841 GGCAATGAAGGCTTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCAAGGGGCGAGCAC 900
QY 925 CCCCACCTCACACCAAGGAGAGACAAATACAGACCCATCAGACCAACCCCGCTTA 984
DB 901 CCCCACCTCACACCAAGGAGAGACAAATACAGACCCATCAGACCAACCCCGCTTA 960
QY 985 CTGTGATGAGTGTGCTGTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCCCCGGAT 1044
DB 961 CTGTGATGAGTGTGCTGTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCCCCGGAT 1020
QY 1045 GCGGAAGGGGATCCCGAAAGCTTCGCTCTCTGTGAGCCGACCCACCTACCCCGAG 1104
DB 1021 GCGGAAGGGGATCCCGAAAGCTTCGCTCTCTGTGAGCCGACCCACCTACCCCGAG 1080
QY 1105 GGTAGTCACTTCGCCCCCGCCCGAGGAGGACCACTGCGAGGCCATTCAACCCAGCTGGTCC 1164
DB 1081 GGTAGTCACTTCGCCCCCGCCCGAGGAGG----- 1109
QY 1165 CTCCAGACAGACCGGGGCGAGCGGAGACTTCCAGAGTTATCTATGGGTGGGTACA 1224
DB 1110 ----- 1109
```

```
QY 1225 CTCCTCAGCCCCCTGAAGCGGGACTTTCCTATTCCTCACCACCTGATGTCCCCAG 1284
DB 1110 ----- 1109
QY 1285 CACTGGTCAATCAGCCACCACTGCTGCCCCCACACAAATGGGCTCAGGATCTCAGGCTTC 1344
DB 1110 -----CACCAGTCCCCCACACAAATGGGCTCAGGATCTCAGGCTTC 1154
QY 1345 CAGTCAAGGCTGACCTTCCGAGCCCTCTGCTGACTTCAAGGCTCGCTCAGTTAGCAT 1404
DB 1155 CAGTCAAGGCTGACCTTCCGAGCCCTCTGCTGACTTCCAGGCTCGCTCAGTTAGCAT 1214
QY 1405 TTCAGTGCATCTACCCACGACGAGTGGGGCCACCCAGAAACAAAGCCCTTCGAA 1464
DB 1215 TTCAGTGCATCTACCCACGAGGTTGGGGCCACCCAGAAACAAAGCCCTTCGAA 1274
QY 1465 ATGATPACTTTTCAATCAGGTTGCTATGGGGCCACCGGACAGGTATGGCCCTTCGCA 1524
DB 1275 ATGATPACTTTTCAATCAGGTTGCTATGGGGCCACCGGACAGGTATGGCCCTTCGCA 1334
QY 1525 GGCTAGGAGGACATCATCACCCAGGAGCCCGAGGATTTAAAGAGCCCTGTGGGGC 1584
DB 1335 GGCTAGGAGGACATCATCACCCAGGAGCCCGAGGATTTAAAGAGCCCTGTGGGGC 1394
QY 1585 AGACAGACATAGCAGGGGTGGGCACTGCTCTCTCTGCAATCTCTAGTCGATTC 1644
DB 1395 AGACAGACATAGCAGGGGTGGGCACTGCTCTCTCTGCAATCTCTAGTCGATTC 1454
QY 1645 TTGCTTTTCTCCGATTCGGGATTTGGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1704
DB 1455 TTGCTTTTCTCCGATTCGGGATTTGGGGGCCACCTCTAAGATGCTCTCTCCAGCCC 1514
QY 1705 TGCTCAACCACTATCCAAATTAGTGCCACCCAGGGGCTGGGACCTCCACATCATCC 1764
DB 1515 TGCTCAACCACTATCCAAATTAGTGCCACCCAGGGGCTGGGACCTCCACATCATCC 1574
QY 1765 ATTGCTTTGCTGCAAGTGCAGATTAACCGGCTGATTGCC 1804
DB 1575 ATTGCTTTGCTGCAAGTGCAGATTAACCGGCTGATTGCC 1614
```

RESULT 5

AAH07100

ID AAH07100 standard; cDNA; 784 BP.

XX AAH07100;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:3935.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 1; SEQ ID NO 395; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 784 BP; 174 A; 224 C; 251 G; 130 T; 0 U; 5 Other;

Query Match 36.5%; Score 661.4; DB 4; Length 784;

Best Local Similarity 98.7%; Pred. No. 1.2e-168;

Matches 687; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 191 GTGCAAGTCTGGTTCGGTTCTGGATTTCGGGCGCGGTCACACGTAGCTGTGCC 250

DB 1 GTGCAAGTCTGGTTCGGTTCTGGATTTCGGGCGCGGTCACACGTAGCTGTGCC 60

QY 251 GGCTCTCTCGGTGAGTCCGCGCGCGGTGCGCGGCGCGGTCAGGTCGCGGGGT 310

DB 61 GGCTCTCTCGGTGAGTCCGCGCGCGGTGCGCGGCGCGGTCAGGTCGCGGGGT 120

QY 311 CCGGGGCGCGGTCAGTTCGGGTCGAGATTGCGGGATCCCGGATGACCGCGGCGCC 370

DB 121 CCGGGGCGCGGTCAGTTCGGGTCGAGATTGCGGGATCCCGGATGACCGCGGCGCC 180

QY 371 CCGGGGCGCGGTCAGTTCGGGTCGAGATTGCGGGATCCCGGATGACCGCGGCGCC 430

DB 181 CCGGGGCGCGGTCAGTTCGGGTCGAGATTGCGGGATCCCGGATGACCGCGGCGCC 240

QY 431 GGATCCCGATGCTTACGAGCAGATGCTCAGCTTTTATAGGTGACCTTACATGTGAC 490

DB 241 GGATCCCGATGCTTACGAGCAGATGCTCAGCTTTTATAGGTGACCTTACATGTGAC 300

QY 491 TTCACTTCAGTTTGTGATCGTAAATGGACAAATTCGAAGCTTCTTACAGTGTGTT 550

DB 301 TTCACTTCAGTTTGTGATCGTAAATGGACAAATTCGAAGCTTCTTACAGTGTGTT 360

QY 551 GAGAGGATTAATGAACAAATGCTTTGAAAGCTTTTTCAGAGGAGGAGCTTCGGAAGCAG 610

DB 361 GAGAGGATTAATGAACAAATGCTTTGAAAGCTTTTTCAGAGGAGGAGCTTCGGAAGCAG 420

QY 611 GGCTTGGCGGCGAGCAGCAGCTGCTTACAGGACCCAGGACGATGAGACCCCG 670

DB 421 GGCTTGGCGGCGAGCAGCAGCTGCTTACAGGACCCAGGACGATGAGACCCCG 480

QY 671 GTGAGAGTGGCGGTCAGTGGGATGCAGACCTTGGGCTTTCAGACCGCTCCCGAGTGGC 730

DB 481 GTGAGAGTGGCGGTCAGTGGGATGCAGACCTTGGGCTTTCAGACCGCTCCCGAGTGGC 540

QY 731 TACCGGCGTCAAGCCAGGACGTCATATGTGGATG-AGACTCTGTTTGGCAGCCAGCAGG 789

DB 541 TACCGGCGTCAAGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCAGCAGG 600

QY 790 CACCGGCGTACCCACCGGACTTCGATCCGCTTGGTGGAGAGGCTACAGACCCAG 849

DB 601 CACCGGCGTACCCACCGGACTTCGATCCGCTTGGTGGAGAGGCTACAGACCCAG 660

QY 850 AGGCGTGGCGAAGGA-GGCATCGAAGGCGCTTGGGG 884

DB 661 AGGCGTGGCGAAGGA-GGCATCGAAGGCGCTTGGGG 696

RESULT 6

AA575219

ID AAS75219 standard; cDNA; 889 BP.

XX AAS75219;

AC AAS75219;

XX DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11023.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2001; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG11032.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 1; SEQ ID NO 11023; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AA594564 represent novel human diagnostic

CC coding sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

Db	361	ACGGGTCCAGACCTGGTGGGAAGAGGTGCGNGGACGGGTCCCTGTAGAGATCCCNATGCCT	420
Qy	445	ACGAGCCAAAGATGCTCAGCTTTATATAGGTGTGACCTACACATGTGACCTTCACTCAGTTTT	504
Db	421	ACGAGCCAAAGATGCTCAGCTTTATATAGGTGTGACCTACACATGTGACCTTCACTCAGTTTT	480
Qy	505	GTGATCCGTAAAATGGACAAATTCGAAGCTACTTTCAC-AGTGCTGTTGAGAGGATTAAT	563
Db	481	GTGATCCGTAAAATGGACAAATTCGAAGCTACTTTCACAGTGTGTTGATAGGATTAAT	540
Qy	564	GAACAATGCTTGTAAAGCTCTTTTCGAGGAGGAGCGCTTCGGAAGCAGGGCTCGGC	619
Db	541	GAACAATGCTTGTAAAGCTCTTTTCGANGAGGAGCGCTTGAAGCAAGGCGCTGGC	596
RESULT 8			
AAH12167/c			
XX	ID	AAH12167 standard; cDNA; 561 BP.	
XX	AC	AAH12167;	
XX	AC		
XX	AC		
DT	26-JUN-2001	(first entry)	
XX	XX	Human cDNA clone (3'-primer) SEQ ID NO:9002.	
XX	DE		
XX	DE		
XX	KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
XX	XX	Homo sapiens.	
XX	OS		
XX	OS		
PN	EP1074617-A2.		
XX	XX		
XX	PD	07-FEB-2001.	
XX	XX		
PF	28-JUL-2000;	2000EP-00116126.	
XX	XX		
PR	29-JUL-1999;	99JP-00248036.	
PR	27-AUG-1999;	99JP-00300253.	
PR	11-JAN-2000;	2000JP-00118776.	
PR	02-MAY-2000;	2000JP-00183767.	
PR	09-JUN-2000;	2000JP-00241899.	
XX	XX		
PA	(HELI-) HELIX RES INST.		
XX	XX		
PI	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, O-suki T;		
XX	XX		
DR	WPI; 2001-318749/34.		

CC	AAH13633	to AAH18742	represent human cDNA sequences; AB992446 to AB958933
CC	represent human amino acid sequences; and AAH13629 to AAH13632	represent	
CC	oligonucleotides, all of which are used in the exemplification of the		
XX	present invention		
XX			
SQ	Sequence 561 BP; 118 A; 142 C; 181 G; 117 T; 0 U; 3 Other;		
	Query Match	27.3%;	Score 494; DB 4; Length 561;
	Best Local Similarity	99.4%;	Fred. No. 2.2e-123;
	Matches 494; Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	1308	CCCCCACACAAATGGGCGCTCAGGATCTCAGGCGCTTCCACGTCAGGGGTGACCTTCGCGA	1367
DB	497	CCCCCCACANAATGGGCGCTCAGGATNTTCAGGCGCTTCCACGTCAGGGGTGACCTTCGCGA	438
QY	1368	GCCTCCCTGGTGACTTCCAGGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCCCACGAC	1427
DB	437	GCCTCCCTGGTGACTTTCAGGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCCCACGAC	378
QY	1428	GAGGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATCATACTCTTTTCATCAGGGTTG	1487
DB	377	GAGGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATACTCTTTTCATCAGGGTTG	318
QY	1488	CCTATGGGGCCACGGCGACAGGTATGGCCCTTTGCCAGGGTAGGAGACATTCATCACCC	1547
DB	317	CCTATGGGGCCACGGCGACAGGTATGGCCCTTTGCCAGGGTAGGAGACATTCATCACCC	258
QY	1548	AGGGAACCCAGGTATTAAAGAGACCCCTGTGGGGCAGACACATAGCAGGGTGGGC	1607
DB	257	AGGGAACCCAGGTATTAAAGAGACCCCTGTGGGGCAGACACATAGCAGGGTGGGC	198
QY	1608	AGTGCCCTCCCTTTATCTCTGACAAATCTCTAGTCGATCTTGCCCTTTTCTCCCGATTGCGG	1667
DB	197	AGTGCCCTCCCTTTATCTCTGACAAATCTCTAGTCGATCTTGCCCTTTTCTCCCGATTGCGG	138
QY	1668	ATTTGGGGGCACCTCTTAAGATGCGCTCTCTCCAGCCCTGTCTCAACATCTCCTCAATTA	1727
DB	137	ATTTGGGGGCACCTCTTAAGATGCGCTCTCTCCAGCCCTGTCTCAACATCTCCTCAATTA	78
QY	1728	GTGCCAACCCAGGGGCGCTGCGACCTCCACATCATCCATTGTCTTGCTGCCAAGTGCAG	1787
DB	77	GTGCCAACCCAGGGGCGCTGCGACCTCCACATCATCCATTGTCTTGCTGCCAAGTGCAG	18
QY	1788	TAAACGCGGTGATTGCC	1804
DB	17	TAAACGCGGTGATTGCC	1
	RESULT 9		
	ACH14455		
ID	ACH14455	standard; cDNA; 477 BP.	
XX	ACH14455;		
AC	ACH14455;		
XX	13-OCT-2003	(first entry)	
DT	13-OCT-2003	(first entry)	
DE	Human adult brain cDNA #1667.		
XX	Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;		
KW	genome mapping; biodiversity; genetic disorder.		
XX	Homo sapiens.		
OS	US2003073623-A1.		
XX	17-APR-2003.		
PD	30-JUL-2001; 2001US-00918995.		
XX	30-JUL-2001; 2001US-00918995.		
PR	(DRMA/) DRMANAC R T.		
XX	(LABA/) LABAT I.		
FA			
PA			

PA (STAC//) STACHE-CRAIN B.
PA (DICK//) DICKSON M C.
XX (JONE//) JONES L W.
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 1667; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 477 BP; 86 A; 174 C; 125 G; 90 T; 0 U; 2 Other;
XX
XX Query Match 24.1%; Score 436.6; DB 8; Length 477;
XX Best Local Similarity 99.1%; Pred. No. 7.4e-108;
XX Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 986 TGTGATGAGTCGTGTTGCTCCCATCTGAAGGGCCAGCTTCGGGGCCCGCGGATG 1045
Db 35 TGGATTCTGTCGTGTTGCTCCCATCTGAAGGGCCAGCTTCGGGGCCCGCGGATG 94
QY 1046 GCGAAGGGGATGCGGCAAGCTCCGTGCTCTCTTGTGACGCCACCACTACCCCGAG 1105
Db 95 GCGAAGGGGATGCGGCAAGCTCCGTGCTCTCTTGTGACGCCACCACTACCCCGAG 154
QY 1106 GGTAGCCACTCGCCCGCCCGGAGGACCACTCGAGCCATTCACCCAGCTGTGCC 1165
Db 155 GGTAGCCACTCGCCCGCCCGGAGGACCACTCGAGCCATTCACCCAGCTGTGCC 214
QY 1166 TCCAAGACAGAGCGGGGCGGCGGAGCTCCAGAAAGTTATCTATGGTGGGTATAC 1225
Db 215 TCCAAGACAGAGCGGGGCGGCGGAGCTCCAGAAAGTTATCTATGGTGGGTATAC 274
QY 1226 TCTTCAGCCCTCGGAGCGGAGCTTCCATTCCTCACCACCTGAATGTCCCGAG 1285
Db 275 TCTTCAGCCCTCGGAGCGGAGCTTCCATTCCTCACCACCTGAATGTCCCGAG 334
QY 1286 ACTGGTTCATCAGCCAGCTGCCCCCCCCACACAAATGGGCTCAGGATCTCAGGCTTCC 1345
Db 335 ACTGGTTCATCAGCCAGCTGCCCCCCCCACACAAATGGGCTCAGGATCTCAGGCTTCC 394
QY 1346 ACCTCAGGGTGACCTTCCGAGGCCCTCGGTGACTTCCAGGGCTCGCTCAGTTAGCA 1405
Db 395 ACCTCAGGGTGACCTTCCGAGGCCCTCGGTGACTTCCAGGGCTCGCTCAGTTAGCA 454
QY 1406 TCAGTGCCATCTACCCCGAGC 1428
Db 455 TCAGTGCCATCTACCCCGAGC 477
XX
XX RESULT 10

ACH43954
ID ACH43954 standard; cDNA; 484 BP.
XX
XX AC ACH43954;
XX
XX DT 13-OCT-2003 (first entry)
XX
XX DE Human foetal brain cDNA #4679.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN US2003073623-A1.
XX
XX PD 17-APR-2003.
XX
XX PF 30-JUL-2001; 2001US-00918995.
XX
XX PR 30-JUL-2001; 2001US-00918995.
XX
XX (DRNA//) DRMANAC R T.
XX (LABA//) LABAT I.
XX (STAC//) STACHE-CRAIN B.
XX (DICK//) DICKSON M C.
XX (JONE//) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 31166; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 484 BP; 118 A; 129 C; 137 G; 85 T; 0 U; 15 Other;
XX
XX Query Match 22.7%; Score 410.6; DB 8; Length 484;
XX Best Local Similarity 93.9%; Pred. No. 8.2e-101;
XX Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
XX
QY 426 CCTGAGATCCGATGCTAGAGCCAGATGCTCAGCTTTATAGTGTGACCTACAT 485
Db 39 CNGNAAANNNNNNNAGNANACNAGCCAGATGCTCAGCTTTATAGTGTGACCTACAT 98
QY 486 GTGACTTTCACCTCAGTTTGTGTGATCCGTAATAATGGCAAAATTCGAAAGCTACTTCACAGTG 545
Db 99 GTGACTTTCACCTCAGTTTGTGTGATCCCAAAATGGCAAAAGGCGAGCTACTTCACAGAG 158
QY 546 CTGTTGAGAGATTAAATGAACAAATGCTTTGTAAGCTCTTTTCAGAGGAGGACCTCGGA 605

Query Match	7.8%;	Score 138;	DB 9;	Length 494;
Best Local Similarity	72.8%;	Pred. No. 5.6e-27;		
Matches 177;	Conservative 0;	Mismatches 66;	Indels 0;	Gaps 0;
QY	933	TCACACCAAGGAAGAAACAATACAGACCCATCAGCCATCAGCCCGCTTACTGTGATG	992	
Db	494	TCACNCAGAGGAGAAAACAATACAGGTTGATTTGGCCACACCCCTTACTGTGATG	435	
QY	993	AGTCGTGTTTGGCTCCCGATCTGAAGGCGCAGCTTCGCGGCCCCCGGGATGCGGAAGG	1052	
Db	434	AGTCACTGTTTGGCTCCCGCCACAGGCGCACGAGAGAGGCTCTCGGACGCCGCTGG	375	
QY	1053	GGGATCGCCCAAGCTCCGTGCTCTTCTGTGACGCGCCACCTACCCCGAGGGGTAGCC	1112	
Db	374	AGGATCGCGCGAAGCTCCGAGCCCTTCTTGACCCACAGCCACCCCTAGGGGAGCC	315	
QY	1113	ACTCGCCCCCGCCAGGAGGACCACTGCGAGCCATTACCCAGCTGGTCCCTCCAGA	1172	
Db	314	ACTCACCTCGCCCCAGGAGAGCCCACTTCGCGGCCATATTTTCCCTAGTCCCAAGCT	255	
QY	1173	CAG 1175		
Db	254	CAG 252		
RESULT 14				
ABN79266				
ID	ABN79266	standard; cDNA; 292 BP.		
XX	AC	ABN79266;		
XX	AC	ABN79266;		
DT	08-JUL-2002	(first entry)		
XX	DE	Human ORF4213 cDNA, SEQ ID NO:8425.		
XX	KW	Human; ORF; open reading frame; ORFX; drug screening; diagnosis;		
KW	KW	disease monitoring; cytokine; cell proliferation; cell differentiation;		
KW	KW	immune modulation; haematopoiesis regulation; tissue growth;		
KW	KW	angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;		
KW	KW	thrombolytic; tumour inhibition; bodily characteristic; fertility;		
KW	KW	behaviour; cancer; proliferative disorder; neurological disorder;		
KW	KW	cardiovascular disease; immune system disorder; organ transplantation;		
KW	KW	tissue growth disorder; tissue regeneration disorder; diabetes mellitus;		
KW	KW	hypothyroidism; cholesterol ester storage disease; infection; vulnery;		
KW	KW	vasotrophic; antiparotetic; antidiabetic; cytostatic; neutropic;		
KW	KW	neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;		
KW	KW	cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;		
KW	KW	dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.		
XX	OS	Homo sapiens.		
XX	PN	WO200190366-A2.		
XX	PD	29-NOV-2001.		
XX	PF	24-MAY-2001; 2001WO-US017076.		
XX	PR	24-MAY-2000; 2000US-0206690P.		
XX	PA	(CURA-) CURAGEN CORP.		
PI	Leach MD, Shinkets RA;			
XX	PI	WPI; 2002-106200/14.		
DR	P-PSDB; ABP35240.			
XX	XX	Novel human polypeptides and polynucleotides useful for diagnosing,		
XX	PT	preventing and treating cardiovascular disease, neurodegenerative,		
PT	PT	hyperproliferative disorders and disorders related to organ		
PT	PT	transplantation.		
XX	XX	Claim 1; Page 2350; 2508pp; English.		
PS	PS			
XX	XX			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 06:05:38 ; Search time 304 Seconds
(without alignments)
3305.974 Million cell updates/sec

Title: US-10-031-589-3
Perfect score: 1811
Sequence: 1 agtgcctcgccctcgccg.....cggctgattgccaaacctgg 1811

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.8	2.7	17493	4	US-09-804-471A-3
C 2	49.8	2.7	17493	3	Sequence 3, Appli
C 3	46.2	2.6	30001	1	US-10-238-709-3
C 4	46.2	2.6	30001	2	Sequence 1, Appli
C 5	45.6	2.5	72928	3	US-08-474-933-1
C 6	43	2.4	4220	1	US-09-009-913-1
C 7	43	2.4	4220	2	US-08-832-883-66
C 8	43	2.4	4220	3	US-08-832-877-66
C 9	43	2.4	13104	3	US-08-256-799-4
C 10	42.4	2.3	505	4	US-08-462-437-4
C 11	41.4	2.3	364	4	US-09-621-976-15639
C 12	41	2.3	202001	4	US-09-621-976-17202
C 13	40.8	2.3	1335	5	US-09-734-674-3
C 14	40.6	2.2	7218	1	PCT-US91-06532-1
C 15	40.4	2.2	7218	2	US-08-232-463-14
C 16	40	2.2	2831	2	US-08-806-713-1
C 17	39.8	2.2	6436	4	US-09-084-669-1
C 18	39.6	2.2	2721	6	US-09-600-099-1
C 19	39.6	2.2	4768	4	US-09-526-193A-16
C 20	39.6	2.2	7218	1	US-08-232-463-14
C 21	39.6	2.2	13987	2	US-08-804-227C-13
C 22	39.6	2.2	43280	2	US-08-804-227C-1
C 23	39.6	2.2	44377	2	US-08-804-227C-7
C 24	39.6	2.2	44377	2	US-08-804-198-1
C 25	39.2	2.2	45546	4	US-09-146-053-6
C 26	39	2.2	527	4	US-09-621-976-2475
C 27	39	2.2	3524	4	US-09-077-940A-3

```

Sequence 3, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 21, Appli
Sequence 24, Appli
Sequence 269, Appli
Sequence 5, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 11, Appli

28 39 2.2 40090 4 US-09-820-004-3
c 29 39 2.2 4403785 3 US-09-103-840A-2
30 38.6 2.1 162450 4 US-09-345-882-1
c 31 38.2 2.1 8438 1 US-07-945-283-1
c 32 38.2 2.1 20284 4 US-09-526-193A-21
c 33 38 2.1 946 4 US-09-904-615-24
c 34 38 2.1 1001 4 US-09-641-638-269
c 35 38 2.1 1881 4 US-09-434-288-5
c 36 38 2.1 48763 4 US-09-916-204-3
c 37 38 2.1 80246 3 US-09-078-294-4
c 38 38 2.1 80595 3 US-09-078-294-3
c 39 38 2.1 174493 4 US-09-804-471A-3
c 40 38 2.1 174493 4 US-10-238-709-3
c 41 37.8 2.1 2263 4 US-09-595-549-1
c 42 37.8 2.1 4257 2 US-08-690-473-1
c 43 37.8 2.1 4257 3 US-09-259-821A-1
c 44 37.8 2.1 4257 3 US-08-843-659-1
c 45 37.8 2.1 12001 1 US-08-458-568A-11

ALIGNMENTS

RESULT 1
US-09-804-471A-3/c
; Sequence 3, Application US/09804471A
; Patent No. 6479269
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO01164
; CURRENT APPLICATION NUMBER: US/09/804,471A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 174493
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(174493)
; OTHER INFORMATION: n = A,T,C or G
US-09-804-471A-3

Query Match 2.7%; Score 49.8; DB 4; Length 174493;
Best Local Similarity 75.8%; Pred. No. 0.0067;
Matches 75; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 494 ACCTCAGTTTGTGATCGTAAATGG--ACAAATTCGAAGCTACTTCACAGTCTGTG 551
Db 91060 ACCTCAGTTTCTCTCAATTGTAATGGCCATAATTACCACCTACTTCAAGAGTCTTG 91001

QY 552 AGAGATTAAATGAACAATGCTTGTAAAGCTTTTGA 590
Db 91000 GAAGGATTAAATGAATAATGATGTAATGCTCTCAGCA 90962

RESULT 2
US-10-238-709-3/c
; Sequence 3, Application US/10238709
; Patent No. 6680188
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO01164DIV
; CURRENT APPLICATION NUMBER: US/10/238,709
; CURRENT FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4

```



```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match          2.6%; Score 46.2; DB 2; Length 30001;
Best Local Similarity 45.9%; Pred. No. 0.029;
Matches 196; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 13 CTTGGCGCGCTAGTACACAGCAGCTGAGTGGACACAGAGACCCCTTCATGTT 72
Db 18498 CCGCGACAGCCGGTCTCTGAAGCCCTCTCGGAGCGCGGAGTTCGCGCTCAACGTGCT 18439

QY 73 TAGGGACCTCCTGGGCTTCAGAGAGCTGGCGCCCGCCCTGGCGGACTCCCGCCATCG 132
Db 18438 GCGCGCGGAGGGAGTTCACGTCCCGCGCTTCGCCCGCCCGTGGAGACCCCTTCC 18379

QY 133 CGGCGCG---CGAATGTCCTGGGTTCGCGTCCGAGTCTCTGCTGCTCTCTGTTGCTG 189
Db 18378 GGGACGTCCCGGTGGAGCGCGGACGCGGTGGGGTCCCGTCACTGTGCGCGGAGCT 18319

QY 190 GGTGCAAGTCTGGGTCTTGGTTTCTGGATTTCGGGGCGCGTTCACAGTAGCCTGTGC 249
Db 18318 CGTCGCGCAGCGAGTGGCGGCTCGCGCGCGGCTCGACAGCGGGGCGACACACCATC 18259

QY 250 CGGCTCTCGGTGAGTCCGCTCGCGCGCGGTGCGCCCGGAGCGGCTAGGCTGCCGGGG 309
Db 18258 GTGATCGGCGGTGTGGCGCGGCGGCGCGCGCGGAGGTGCCAGCCGCTGATGT 18199

QY 310 TCCGGGGCCCGAGGATTCGGGCTCAGATTGACGGGGATCCGGATGACGCGCGCGCC 369
Db 18198 ACTGGCGCGCTCTACGCGCGTGGCGGTGGAGGAGATCCGGAGCGCGCGCTGA 18139

QY 370 CCGCGCCCTCACCGACGGTCCAGACCTGTTGGGAGAGTGGCGGAGCGGTCCTG 429
Db 18138 CCTGGCGCGGAGGCTGAGCCCGCGCGGTGGCGCGGAGCGGCGGCGCGGCGG 18079

QY 430 AGGATCC 436
Db 18078 AGCAGCC 18072

RESULT 5
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: AxyS Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match          2.5%; Score 45.6; DB 3; Length 72928;
Best Local Similarity 62.1%; Pred. No. 0.064;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 471 GTGTGACCTACATGTGACTTACCTCAGTTTTTGTGATCCGTAAATGACAAATTGCA 530
Db 66880 GAGTGAGTTATATAAGCTTAGCATCTAAGTTTTTCTCATCTGGAAAATGGAGTTAA 66939

QY 531 AGCTACTTCACAGTCTGTTGAGAGGATTAATGAAAACAATGCTTGTAAAGCTCTT 586
Db 66940 ATCTACTGCATTGGGCTGTTGTAAGATTAAATTAACAAGAATGTGAAAGCACT 66995

RESULT 6
US-08-832-883-66
; Sequence 66, Application US/08832883
; Patent No. 5807881
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; APPLICANT: Baldi, Alphonso
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-66

Query Match          2.4%; Score 43; DB 1; Length 4220;
Best Local Similarity 73.3%; Pred. No. 0.09;
```

Matches 55; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 Qy 497 TCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAGTGTCTGTGAGAG 556
 Db 1130 TCAGTATGGTGATCTCTAAACTGGAGATATTGTGTTTACCTCACAGAGCTGTCTGAAG 1189
 Qy 557 ATTAATGAAACAAT 571
 Db 1190 ATTAATAAGGCAAT 1204

RESULT 7
 US-08-832-877-66
 ; Sequence 66, Application US/08832877
 ; Patent No. 5840506
 ; GENERAL INFORMATION:
 ; APPLICANT: Giordano, Antonio
 ; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
 ; NUMBER OF SEQUENCES: 116
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
 ; STREET: Suite 1800 Two Penn Center Plaza
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/832,877
 ; FILING DATE:
 ; CLASSIFICATION: 436
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Monaco, Daniel A
 ; REGISTRATION NUMBER: 30,480
 ; REFERENCE/DOCKET NUMBER: 8321-13 US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 568-8383
 ; TELEFAX: (215) 568-5549
 ; INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4220 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-832-877-66

Query Match 2.4%; Score 43; DB 2; Length 4220;
 Best Local Similarity 73.3%; Pred. No. 0.09; Indels 0; Gaps 0;
 Matches 55; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 Qy 497 TCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAGTGTCTGTGAGAG 556
 Db 1130 TCAGTATGGTGATCTCTAAACTGGAGATATTGTGTTTACCTCACAGAGCTGTCTGAAG 1189
 Qy 557 ATTAATGAAACAAT 571
 Db 1190 ATTAATAAGGCAAT 1204

RESULT 8
 US-08-256-799-4/c
 ; Sequence 4, Application US/08256799
 ; Patent No. 622094
 ; GENERAL INFORMATION:
 ; APPLICANT: HANSSON, Lemnart
 ; APPLICANT: STROEMQVIST, Mats
 ; APPLICANT: BERGSTROEM, Sven

; APPLICANT: HERNELL, Olle
 ; APPLICANT: Toernell, Jan
 ; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR
 ; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/256,799
 ; FILING DATE: 06-DEC-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: DK 88/92
 ; FILING DATE: 23-JAN-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: COOPER, Iver P.
 ; REGISTRATION NUMBER: 28,005
 ; REFERENCE/DOCKET NUMBER: HANSSON-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 13104 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: ps459
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 1..8834
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 8868..10014
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 10511..12277
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: 8835..8867
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: 10015..10510
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: 12278..12443
 ; US-08-256-799-4

Query Match 2.4%; Score 43; DB 3; Length 13104;
 Best Local Similarity 62.6%; Pred. No. 0.15;
 Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
 Qy 484 ATGTGACTTCACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAG 543
 Db 4542 ATCTTCAGCACCTCAGTTTCATCCCTGTAGATGAGGATACCTAATTCCTCCCTTATGG 4483
 Qy 544 TGCTGTTGAGAGGATTAAATGAACAATGCTTGTAAAGCTCTTTGCA 590
 Db 4482 GAGTGTGAAAGGATGAAGTGAAGAAATCCATGTTAACTTCTTAGCA 4436

RESULT 9
US-08-462-437-4/c
; Sequence 4, Application US/08462437
; Patent No. 6232094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROM, Sven
; APPLICANT: HERNELL, Olie
; APPLICANT: TOERNELL, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,437
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: ps459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; FEATURE:
; NAME/KEY: intron
; LOCATION: 10511..12277
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8935..8867
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10015..10510
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12278..12443
US-08-462-437-4

Query Match

2.4%; Score 43; DB 3; Length 13104;

Best Local Similarity 62.6%; Pred. No. 0.15;
Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 484 ATGTGACTTCACCTCAGTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAG 543
Db 4542 ATCTTCAGCACCTCAGTTTCATCCCTCTGTAGATGAGGATATAATTCCTCCCTTATGG 4483
Qy 544 TGCTGTTGAGAGGATTAAATGAAACAATGCTTCTAAAGCTCTTTGCA 590
Db 4482 GAGTGTGAAAGGATGAAGTGAGAAATCCATGTTAACTTCTTAGCA 4436
RESULT 10
US-09-621-976-15639/c
; Sequence 15639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976.
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15639
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15639

Query Match 2.3%; Score 42.4; DB 4; Length 505;
Best Local Similarity 10.8%; Pred. No. 0.05;
Matches 37; Conservative 158; Mismatches 149; Indels 0; Gaps 0;
Qy 1017 AAGCGCCAGCTTCGGGGCCCGGGATGCGGAGGGGGATGCGCAAAAGTCCGTGTC 1076
Db 370 AAARRRMAAGGSGYCGWTSYTSKMTGRKSWTKRMMWTYSGMWTYKCTTKTKY 311
Qy 1077 TCTTGTGAGCCACCACTACCCAGGGGTAGCCACTCGCCGCCGCCAGGAGGCAC 1136
Db 310 TGWKSXKTRWTCSTWRKYMMWMSGWARSMKSWARSWYMWACWCMMSASAYRARRSMYG 251
Qy 1137 CACTGCGAGCCATTCACCCAGCTGTCCTCCCAAGACAGAGCCGGGCCAGCGCAGACT 1196
Db 250 ARSMRTRAGAWWRARRGKKRARRKSSMRKSKMSMRKSWARSWYMWACWCMMSASAYRARRSMYG 191
Qy 1197 CCCAGAGTTATCTATGGTGGTTACACTTTCCAGCCGCCCTTCAAGCGGGAGCTTTCC 1256
Db 190 MGSKMSCRGTCAKMWRYARYAKYVASSMGKYMGRWCYAKCARMYGYRSTRSGSR 131
Qy 1257 ATTCCTCACCACCTGAATGTCGCCAGACTGTGTCATCCAGCCACAGTGCCTCCGCCACA 1316
Db 130 GMKYRRRRKYMMKYMMWMSWYCYRMGAAYGMSARAYRYMYSASACACKMSRMKMSWMSM 71
Qy 1317 CAAATGGGCGCTCAGATCTCAGGCGCTCCACGTCAGGGGTGACC 1360
Db 70 WWRCSRYSYRCWMSGKWCYSCCGYCCSACRMCYCTWRMRKMSWYS 27

RESULT 11
US-09-621-976-17202/c
; Sequence 17202, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 17202
 ; LENGTH: 364
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-621-976-17202

Query Match 2.3%; Score 41.4; DB 4; Length 364;
 Best Local Similarity 17.1%; Pred. No. 0.08;
 Matches 51; Conservative 116; Mismatches 132; Indels 0; Gaps 0;
 QY 91 CAGGAGCGTGGCGCCCGCCCTGGGGGAGTCCCGCCATCCGGCGCGGAATGGTCGG 150
 DB 328 CTGCGACCTGCTGCTGCTGGCGGSCCWGGRGSMCHYSRKGMSCMYSTGRSMVK 269
 QY 151 GTGCGTCCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210
 DB 268 STYKRWRSRAGSMWTGYRMSKMTGSTRSCTSKKRGKSTSKYASTSGKSKYMTCT 209
 QY 211 GGTTCCTGAGTTCGGCGCGGCTTCACAGGTAGCTGTGCGGCTCCTCGGTGAGTCGT 270
 DB 208 KSKKKRYSATYYSCMKWKYCMMSATYSGCMRWTCYSCMSRYSCTSYSRKCSK 149
 QY 271 CCGCGCGCGTGGCCCGGAGCGCTAGCTGCGCGGGTCCGGGCGCCCGCAGGCAATCG 330
 DB 148 TGMKGGKCYRMYRGRMYRMSRGARYTKSRGCRSTKRYRTKTCASWGAKTWCC 89
 QY 331 GGCTGCAGATTGACGGGGATCCCGGATGACCGCGGCGCCCGCCCTCACCGAGGG 389
 DB 88 MRMGSTGASTMRKSKYRKMWSKMKGYGYSWYSMSGSGCKSGKSSYCRSYSS 30

RESULT 12

US-09-734-674-3
 ; Sequence 3, Application US/09734674
 ; Patent No. 6498022
 ; GENERAL INFORMATION:
 ; APPLICANT: WEI, Ming-Hui et al
 ; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 ; FILE REFERENCE: CL001018
 ; CURRENT APPLICATION NUMBER: US/09/734,674
 ; CURRENT FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 202001
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(202001)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-734-674-3

Query Match 2.3%; Score 41; DB 4; Length 202001;
 Best Local Similarity 63.9%; Pred. No. 1.9;
 Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 494 ACCTCAATTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAGTGTCTTTCAG 553
 DB 29199 ACCTCAATTTTGTGATCCGTAATAATGACAAATTCGAAGCTACTTCACAGTGTCTTTCAG 553
 QY 554 AGATTAAATGAACAAATCTTGTAAAGCTTTTGA 590
 DB 29259 AAGATTAAATGAACAAATCTTGTAAAGCTTTTGA 590

RESULT 13

PCR-US91-06532-1/c
 ; Sequence 1, Application PC/TUS9106532

; GENERAL INFORMATION:
 ; APPLICANT: Roizman, Bernard
 ; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
 ; TITLE OF INVENTION: Vaccines and Methods
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
 ; ADDRESSEE: Bicknell,
 ; STREET: Two First National Plaza Suite 2100
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60603
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US91/06532
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gruber, Lewis S.
 ; REGISTRATION NUMBER: 30,060
 ; REFERENCE/DOCKET NUMBER: 27373/8235
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/346-5750
 ; TELEFAX: 312/984-9740
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1335 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 PCT-US91-06532-1

Query Match 2.3%; Score 40.8; DB 5; Length 1335;
 Best Local Similarity 44.8%; Pred. No. 0.21;
 Matches 156; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
 QY 82 CTGTGGGCTCAGGAGGTGCGCCCGCCCTGGGGCGGACTCCCCCATCGCGGGCGCGA 141
 DB 754 CCCAGACACACAGGTGGCGCACCCCGGACGTGGGGCGAGAAGCGCACCCGCGGGGTGCG 695
 QY 142 ATGTCGGGTGCGTCCGCGAGTGTGCTGGTGTCTCCCTGTTGCTGGTGTGCAAGTGC 201
 DB 694 CGGGGTGCGGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGG 635
 QY 202 TGGGTTCTGGGTTTCTGGAATTCGCGGGCGCTTCACACGTAGCTGTGCGGGTCTCTCGGG 261
 DB 634 TCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGG 575
 QY 262 TGAATCGTCCGCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGG 321
 DB 574 GGGCGAGGCGCGCAGGTGCTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGGGTTCGCGGG 515
 QY 322 GGCATTCCGGGCTGCGAGATTGACGGGGATCCCGGATGACACCGCGCCCGCCCGCCCTCA 381
 DB 514 GCGCGAAGGCGCGTGAAGGGGGTGGAGGGGTAGCCCGCCCGCCCGCCCGCCCGCGCG 455
 QY 382 CCAGCGGTCCAGACCTGTTGGGAAGAAGTGGCGGGACGGGTCCCTG 429
 DB 454 GCGGTGGGAGACCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGG 407

RESULT 14

US-08-232-463-14
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367
 ; GENERAL INFORMATION:

This Page Blank (uspto)

GenCore version 5.1.5
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 05:02:45 ; Search time 4878 Seconds
(without alignments)
11086.595 Million cell updates/sec

Title: US-10-031-589-3
Perfect score: 1811
Sequence: 1 agtgcctggccctcgccg.....cggcgtattgccaacctgg 1811

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1012.2	55.9	1152	13	BX358473
2	976.8	53.9	1201	13	BX339727
3	972	53.7	1201	13	BX339726
4	932.4	51.5	1196	9	AL523765

5	930.4	51.4	1201	13	BX403928
6	924.4	51.0	1201	9	AL560267
7	910.4	50.3	979	13	BX354462
8	898.6	49.6	933	13	BX345478
9	889.2	49.1	1030	9	AL582549
10	884.2	48.8	1047	12	BM903591
11	882.2	48.7	1201	9	AL582286
12	880.6	48.6	1201	9	AL517881
13	876.8	48.4	1201	13	BX384471
14	876.6	48.4	1201	13	BX376800
15	863.6	47.7	1201	13	BX332260
16	860.2	47.5	1017	12	BM921038
17	857.6	47.4	1048	9	AL561282
18	852	47.0	1201	9	AL560832
19	850.6	47.0	1201	13	BX394189
20	849	46.9	1201	13	BX353787
21	835.4	46.1	888	13	BX346001
22	830.2	45.8	934	13	BX352694
23	828	45.7	1201	9	AL581746
24	826.8	45.7	950	13	BQ943060
25	824	45.5	890	13	BX366971
26	817.6	45.1	1201	9	AL516514
27	812.4	44.9	1067	9	AL555180
28	803.6	44.7	879	13	BU526901
29	803.4	44.4	907	13	BQ962573
30	791.6	43.7	1042	13	BX366455
31	785.8	43.4	1071	13	BX356915
32	781.2	43.1	867	12	B1758555
33	778.4	43.0	848	13	BU540911
34	778.2	43.0	1029	13	BQ057743
35	778.2	43.0	1079	13	BQ278768
36	777.6	42.9	977	12	BM468165
37	766.6	42.3	964	13	BQ670315
38	765.2	42.3	841	12	B1858698
39	764	42.2	1020	12	BM559325
40	763.4	42.2	990	12	BM911750
41	761.2	42.0	799	12	B1761620
42	760	42.0	991	13	BX375937
43	758.4	41.9	849	12	EG281865
44	758.4	41.9	909	14	CF995108
45	753.8	41.6	1197	13	BX384470

ALIGNMENTS

RESULT 1
BX358473
LOCUS BX358473 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION Clone CS0DI039YB15 5-PRIME, mRNA sequence.
ACCESSION BX358473
VERSION BX358473.1 GI:30347985
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 1152)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
CONTACT Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6579.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI039CA08QP1&cluster=6579.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitroGen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0D1039CA08Q01.

```

FEATURES
  source
    1..1152
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS0D1039YB15"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-oligo(dT)
    primer. Five prime end enriched, double-strand cDNA was
    digested with Not I and cloned into the Not I and EcoR V
    sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
  Query Match      55.9%; Score 1012.2; DB 13; Length 1152;
  Best Local Similarity 98.4%; Pred. No. 2.1e-232;
  Matches 1060; Conservative 4; Mismatches 8; Indels 5; Gaps 4;

Qy 196 AAGTCTGGTTCCTGGTTCTTGGATTTCGGGGCCGTTTCACAGTAGCCTGTGCCGGCTC 255
Db 71 AAGTCTGGTTCCTGGTTCTTGGATTTCGGGGCCGTTTCACAGTAGCCTGTGCCGGCTC 130
Qy 256 CTCGGGTGAGTCCGTCGGCGGGGTGTCCTCCGGAGCGGCTAGGCTGCCGGGGGTCCGGG 315
Db 131 CTCGGGTGAGTCCGTCGGCGGGGTGTCCTCCGGAGCGGCTAGGCTGCCGGGGGTCCGGG 190
Qy 316 GCCCCAGGCATTCGGGGCTGCAGATTGACGGGGATCCGGATGACACCGGGCCCGCGG 375
Db 191 GCCCCAGGCATTCGGGGCTGCAGATTGAC-GGGATCCGGATGACACCGGGCCCGCGG 249
Qy 376 CCTCACCGACGGTTCAGACTGTGTGGGAAGAGTGGCGGACCGGGTCCCTGAGGATC 435
Db 250 CCTCACCGACGGTTCAGACTGTGTGGGAAGAGTGGCGGACCGGGTCCCTGAGGATC 309
Qy 436 CCGATGCTACGACGACCAAGATGCTCAGCTTTATAGTGTGACCTACACATGTGACTTAC 495
Db 310 CCGATGCTACGACGACCAAGATGCTCAGCTTTATAGTGTGACCTACACATGTGACTTAC 369
Qy 496 CTCAGTTTCTGTATCCGTAATAATGGAATAATTCGAAGTACTTTCACAGTCTGTGTGAGAG 555
Db 370 CTCAGTTTCTGTATCCGTAATAATGGAATAATTCGAAGTACTTTCACAGTCTGTGTGAGAG 429
Qy 556 GATTAATGAAACATGCTTTGTAAGCTCTTTTCAGAGGAGGACCTTCGGAACGAGGGCT 615
Db 430 GATTAATGAAACATGCTTTGTAAGCTCTTTTCAGAGGAGGACCTTCGGAACGAGGGCT 489
Qy 616 GGCCGGCAGACACACCTGCTGCACGAGGACACAGGACGATGAAGACCCCGGTGA 675
Db 490 GGCCGGCAGACACACCTGCTGCACGAGGACACAGGACGATGAAGACCCCGGTGA 549
Qy 676 GCTGGCGGTGAGTGGATGAGACCTCGGCTTCAGACCGGTGCGAGGTGGCTACCG 735
Db 550 GCTGGCGGTGAGTGGATGAGACCTCGGCTTCAGACCGGTGCGAGGTGGCTACCG 609
Qy 736 GGTCAAGGCCAGGACGCTATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCGG 795
Db 610 GGTCAAGGCCAGGACGCTATATGTGATGAGACTCTGTTTGGCAGCCAGCAGGACCGG 669
Qy 796 GCTTACCCACCGGACTTCGATCCGCTCGGTGGAGAAAGGCTTAACAGAACCCAGAGGGCT 855
Db 670 GCTTACCCACCGGACTTCGATCCGCTCGGTGGAGAAAGGCTTAACAGAACCCAGAGGGCT 729
Qy 856 GGGCAAGGAGGACTCGAAGGCTTTGGGGCAAGGGGAGCTGTGAGACCCACCCCTCAAG 915
Db 730 GGGCAAGGAGGACTCGAAGGCTTTGGGGCAAGGGGAGCTGTGAGACCCACCCCTCAAG 789
Qy 916 GGGCAGCAGCCCGACCCCTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACAC 975
Db 790 GGGCAGCAGCCCGACCCCTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACAC 849
Qy 976 CCGTCTTACTGTGATGAGTCTGTTTGGCTCCCGATCTGAAGGCGGCAGCTTCGGGGC 1035
Db 129 CCGTCTTACTGTGATGAGTCTGTTTGGCTCCCGATCTGAAGGCGGCAGCTTCGAGGCGCTTGG 188

```

```

Db 850 CCGGTCTTACTGTGATGAGTCTGTTTGGCTCCCGATCTGAAGGCGGCACCTTTCGGGG 909
Qy 1036 CCGCGGATGCGGAAGGGGATGCGCAAGACTCGTCTCTTGTGGAGCCACACACC 1095
Db 910 CCGCGGATGCGGAAGGGGATG-CGCAAGCTCCGCTCTCTTGTGGAGCCACACACC 968
Qy 1096 TACCCCGAGGGTAGCCTC- GCCCGCCCGCAGGGAGGAGCAGCAGCCAGCCATTCAC 1153
Db 969 TACCCCGAGGGTAGCCTCCGCCCCCGCCCAAGAGGAGGACCACTTGGAGCCATTCAC 1028
Qy 1154 CCAGTGTCTCCTCAAGACAGAGCCGGGCGCAGAGCTCCAGAGAGTTATCTATG 1213
Db 1029 CCAGTGTCTCCTCAAGACAGAGCCGGGSCAGCGGAGACTCCAGAGAGTTATCTATD 1088
Qy 1214 GGTGGTTTACACTCTTCCAGCCCTTCAAGGGGAGCTTTCCTATCCCTCAGCCAC 1270
Db 1089 GGTGGTTTACACTCTTCCAGCCCTTGAAGCGGGAC-TTTCATTCCTCAGCCAC 1144

RESULT 2
BX339727
LOCUS BX339727 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1076YG21 5-PRIME, mRNA sequence.
ACCESSION BX339727
VERSION BX339727.1 GI:30335832
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li.W.B., Gruber.C., Jesse.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6579.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1076AD11Q1&cluster=6579.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1076AD11Q1.

FEATURES
  source
    1..1201
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS0D1076YG21"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-oligo(dT)
    primer. Five prime end enriched, double-strand cDNA was
    digested with Not I and cloned into the Not I and EcoR V
    sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
  Query Match      53.9%; Score 976.8; DB 13; Length 1201;
  Best Local Similarity 99.0%; Pred. No. 7.2e-224;
  Matches 1008; Conservative 5; Mismatches 2; Indels 3; Gaps 3;

Qy 762 ATGAGACTCTGTTTGGCAGCCCGAGGAGCACCCTACCCACCGGACTTCGATCCG 821
Db 69 ATGAGACTCTGTTTGGCAGCCCGAGGAGCACCCTACCCACCGGACTTCGATCCG 128
Qy 822 CCTGGGTGGAGAGGCTTAACAGAACCCAGAGGGGTGGGCAAGGAGGATCGAAGGCTTGG 881
Db 129 CCTGGGTGGAGAGGCTTAACAGAACCCAGAGGGGTGGGCAAGGAGGATCGAAGGCTTGG 188

```



```

RESULT 5
BX403928      1201 bp      mRNA      linear      EST 15-MAY-2003
LOCUS        BX403928 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
DEFINITION   CLOB006ZC02 5-PRIME, mRNA sequence.
ACCESSION    BX403928
VERSION      BX403928
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 1201)
JOURNAL      Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
COMMENT      Full-length cDNA libraries and normalization
              Unpublished (2001)
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. This sequence belongs to sequence cluster 6579.r For
              more information about this cluster, see
              http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CLOB006ZC02RP1&cluster=6579.r. Contact :
              Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/Invitrogen Corporation 1600
              Paraday Avenue Genoscope sequence ID : CLOB006ZC02RP1.
              Location/Qualifiers
                1. 1201
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="CLOB006ZC02"
                  /cissue_type="NEUROBLASTOMA"
                  /clone_lib="Homo sapiens NEUROBLASTOMA"
                  /notes="vector: pCMVSPORT-6; 1st strand cDNA was primed
                  with a NotI-oligo(dT) primer. Five prime end enriched,
                  double-strand cDNA was digested with Not I and cloned into
                  the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                  Library was not normalized."

FEATURES
source
Query Match      51.4%; Score 930.4; DB 13; Length 1201;
Best Local Similarity 96.3%; Pred. No. 1.1e-212;
Matches 1019; Conservative 9; Mismatches 15; Indels 15; Gaps 7;

QY 213 TTCTGATTCGCGGCGGTTACACGTAGCCTGTCCGCGCTCCTCGGGTGTGTCGTC 272
DB |||||
DB 36 TTCTGATTCGCGGCGGTTACACGTAGCCTGTCCGCGCTCCTCGGGTGTGTCGTC 95
QY 273 GCGGCGGGTCCCGCGGACGCGCTAGGCTGCGCGGGGTCCGGGCGCCAGGCATTCCGGG 332
DB |||||
DB 96 GCGGCGGGTCCCGCGGACGCGCTAGGCTGCGCGGGGTCCGGGCGCCAGGCATTCCGGG 155
QY 333 CTGAGATTGACGGGATCCCGATGACCGCGCGCGCGCGCGCTCACCAGCGGTCC 392
DB |||||
DB 156 CTGAGATTGACGGGATCCCGATGACCGCGCGCGCGCGCGCTCACCAGCGGTCC 215
QY 393 AGACCTGTTGGGAAGAGTGGCGGACGGGTCCCTGAGGATCCCGATCGCTACGAGCCA 452
DB |||||
DB 216 AGACCTGTTGGGAAGAGTGGCGGACGGGTCCCTGAGGATCCCGATCGCTACGAGCCA 275
QY 453 AGATGCTCAGCTTTATAGTGTGACCTACACATGTGACCTTACCTCAGTTTGTGATCG 512
DB |||||
DB 276 AGATGCTC-----AGGTGTGACCTACACATGTGACCTTACCTCAGTTTGTGATCG 327
QY 513 TAAATGGAACAATTCGAAGCTACTTCACAGTCTGTGTGAGAGATTAAATGAACAATG 572
DB |||||
DB 328 TAAATGGAACAATTCGAAGCTACTTCACAGTCTGTGTGAGAGATTAAATGAACAATG 387
QY 573 CTGTGAAGCTCTTTGAGGAGGAGCCTCGGAAGCAGGCGCTGCGCGGAGCACACC 632
DB |||||

```

Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
 Faraday Avenue Genoscope, sequence ID : CS0DG007BE09QPl.

FEATURES

source

1. 1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DG007YI18"
 /tissue_type="B CELLS (RAMOS CELL LINE)"
 /cell_line="RAMOS CELL LINE"
 /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
 /notes="vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 51.0%; Score 924.4; DB 9; Length 1201;
 Best Local Similarity 95.6%; Pred. No. 3e-211;
 Matches 1000; Conservative 4; Mismatches 35; Indels 7; Gaps 5;

```

Qy 15 TCAGCGCCCTAGTACACACGACCTGAGTGGTGGACGAGGACCCCTCCATGTTTA 74
Dy 60 TCAGCGCCCTAGTACACACGACCTGAGTGGTGGACGAGGACCCCTCCATGTTTA 119
Qy 75 GGGACCTCCCTGGGCTCAGAGAGGCTGGCGCCCGCCCTGGCGGAGCTCCCGCCATCCGGC 134
Dy 120 GGGACCTCCCTGGGCTCAGAGAGGCTGGCGCCCGCCCTGGCGGAGCTCCCGCCATCCGGC 179
Qy 135 GGGCGGAATCGTCGGGTCCGCTCCGAGTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTC 194
Dy 180 GGGCGGAATCGTCGGGTCCGCTCCGAGTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTCTGCTGTC 239
Qy 195 AAAGTGTCTGGTCTGGTCTTGGATTCCGCGGCGCTTCCACAGTGTCTGCTGTCTGCTGTCTGCTGTCTG 254
Dy 240 AAAGTGTCTGGTCTGGTCTTGGATTCCGCGGCGCTTCCACAGTGTCTGCTGTCTGCTGTCTGCTGTCTG 299
Qy 255 CTTGGGTGAGTCTCGTCCGCGCGCTGCTCCCGGACGCGCTAGGTGTCGGGGGTCCGG 314
Dy 300 CTTGGGTGAGTCTCGTCCGCGCGCTGCTCCCGGACGCGCTAGGTGTCGGGGGTCCGG 359
Qy 315 GGGCCAGGATTCGGGCTGCGATTCAGGGGATCCCGGATGACCGCGCGCGCGCGCGCGCGCGCGCGCG 374
Dy 360 GGGCCAGGATTCGGGCTGCGATTCAGGGGATCCCGGATGACCGCGCGCGCGCGCGCGCGCGCGCGCG 419
Qy 375 GCGCTCACCGAGCGGTCCAGACCTGCTGGGAGAGAGTGGGGGACGGGTCCCTCGAGGAT 434
Dy 420 GCGCTCACCGAGCGGTCCAGACCTGCTGGGAGAGAGTGGGGGACGGGTCCCTCGAGGAT 479
Qy 435 CCGATGCTTACAGCGGATGCTCAGCTTTATAGTGTGACCTACATGTGACTTCA 494
Dy 480 CCGATGCTTACAGCGGATGCTCAGCTTTATAGTGTGACCTACATGTGACTTCA 539
Qy 495 CTTGAGTTTGTGATCCGTAATGACAAATTCGAAGTACTTTACAGTGTGTTGAGA 554
Dy 540 CTTGAGTTTGTGATCCGTAATGACAAATTCGAAGTACTTTACAGTGTGTTGAGA 599
Qy 555 GGATTAATGAACAATGCTTTGTAAGCTTTTTCAGGAGGAGGCTCGGAAGAGGCGCC 614
Dy 600 GGATTAATGAACAATGCTTTGTAAGCTTTTTCAGGAGGAGGCTCGGAAGAGGCGCC 659
Qy 615 TGGCCGCGAGAGCACCTGCTGCTACAGGAGCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 674
Dy 660 TGGCCGCGAGAGCACCTGCTGCTACAGGAGCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 719
Qy 675 AGTGCGCGCTGAGTGGATGACAGACCTCGGCTTTCAGACCGCTGCGGAGTGTCTACC 734
Dy 720 AGTGCGCGCTGAGTGGATGACAGACCTCGGCTTTCAGACCGCTGCGGAGTGTCTACC 779
Qy 735 GGGTCAGGCGGAGGAGTCTATGTGGATGAGACTCTGTTTGGAGCGCCAGGAGGAGGAGGAGGAGGAGGAG 794

```

```

Db 780 GGGTCAAGGCCAGGACGTCATATGTGGATGAGACTCTGTTTGGCAGCCAGCAGGACCC 839
Qy 795 GGCCTACCCACCGGACTTCGATCGGCGCTGGGTGGAGAGGCTTAAACAGAACGAGGCG 854
Db 840 GGCCTACCCACCGGACTTCGATCGGCGCTGGGTGGAGAGGCTTAAACAGAACGAGGCG 899
Qy 855 TGGCAAGGAGGAGCATCGAAGGCTTGGGGCAAGAGGAGGAGCTGTGAGACACCCCTCAA 914
Db 900 TGGCAAGGAGGAGCATCGAAGGCTTGGGGCAAGAGGAGGAGCTGTGAGACCA-CCCCTCAA 958
Qy 915 GGGCAGACCCCCACCCCTCACCAAGGAGAGAGAAACAAATACAGACCCATAGCCACA 974
Db 959 GGGCAGCA-CCCCACCCCTCACCAAGGAGAGAGAAATACAGA---CCATAGCCACA 1014
Qy 975 CCGCTTCTTCTGATGAGTGTGCTTGGCTCCCGATCTGAGAGGCGCGAGCTTCGGGG 1034
Db 1015 -CCGCTTCTTCTGATGA-TCGCTGTTGGCTCCCGATCTGTAGGCGGACTTCGGGGCC 1072
Qy 1035 CCGCGCGATGCGAAGGGGAGGCC 1060
Db 1073 CCGATGGGAGGGGATGCGCAATC 1098

```

RESULT 7
 BX354462/c
 LOCUS
 DEFINITION BX354462 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DC020YC08 3-PRIME, mRNA sequence.
 ACCESSION BX354462
 VERSION BX354462.1 GI:30383876
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 979)
 Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6579.r. For
 more information about this cluster, see
<http://www.genoscope.cns.fr/>
 cgi-bin/cluster.cgi?seq=CS0DC020BB04NP1&cluster=6579.r. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
 Faraday Avenue Genoscope, sequence ID : CS0DC020BB04NP1.

FEATURES

source

1. 979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DC020YC08"
 /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoRV
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 50.3%; Score 910.4; DB 13; Length 979;
 Best Local Similarity 96.8%; Pred. No. 6.3e-208;
 Matches 953; Conservative 14; Mismatches 12; Indels 5; Gaps 4;
 Qy 813 TCGATCCGCTGGGTGAGAGGCTTAAACAGACGAGGCGTGGGCAAGGAGGAGGATCGA 872
 Db 979 TCGATCCGCTGGTGTG--AGAGCTTACAGACGAGGCGT-GGSAAGGAGGATCGA 923

```
Qy 873 AGGCTTTGGGGCAAGAGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCACCCCCACCC 932
Db 922 AGGCTTT-GGGGCAAGAGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCACCCCCACCC 864
Qy 933 TCACACCAAGAGAGAGAAACAAATACAGACCCATCAGCACACCCCGTCTTACTGTGATG 992
Db 863 TCACACCAAGAGAGAGAAACAAATACAGACCCATCAGCACACCCCGTCTTACTGTGATG 804
Qy 993 AGTGGCTTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCGCCCGCGATGGCGAAGG 1052
Db 803 AGTGGCTTTGGCTCCCGATCTGAAGGCGCCAGCTTCGGGGCGCCCGCGATGGCGAAGG 744
Qy 1053 GGGATCGCGCAAGCTCCGCTCTCTTGTGAGCGCCACACCATCACCAGGGGTAGCC 1112
Db 743 GGGATCGCGCAAGCTCCGCTCTCTTGTGAGCGCCACACCATCACCAGGGGTAGCC 684
Qy 1113 ACTCGCCCCCCCCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGTGCTCCCTCAAGA 1172
Db 683 ACTCGCCCCCCCCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGTGCTCCCTCAAGA 624
Qy 1173 CAGAGCGGGGCGAGCGGAGAGCTCCAGAGTTATCTATGGTGGGTACACTTTTCAC 1232
Db 623 CAGAGCGGGGCGAGCGGAGAGCTCCAGAGTTATCTATGGTGGGTACACTTTTCAC 564
Qy 1233 GCCCCTGAAGCGGGACTTTCCATTTCCCTCACCACCTGAATGTCCCGAGCACTGGTC 1292
Db 563 GCCCCTGAAGCGGGACTTTCCATTTCCCTCACCACCTGAATGTCCCGAGCACTGGTC 504
Qy 1293 ATCAGGCCACAGTGGCCCCCACCACAAATAGGCGCTCAGGATCTCAGGCTTCCAGTCAG 1352
Db 503 ATCAGGCCACAGTGGCCCCCACCACAAATAGGCGCTCAGGATCTCAGGCTTCCAGTCAG 444
Qy 1353 GGGTGACTTCCGGAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTACGATTTAGTCG 1412
Db 443 GGGTGACTTCCGGAGCCCCCTGGTGACTTCCAGGGCTCGCTCAGTACGATTTAGTCG 384
Qy 1413 CATCTACCCACAGCAGAGTGGGGCCACCCAGAGAACCAAGACCCCTTGGAAATGATACT 1472
Db 383 CATCTACCCACAGCAGAGTGGGGCCACCCAGAGAACCAAGACCCCTTGGAAATGATACT 324
Qy 1473 CTTTCATCAGGTTGCTATGGGGCCACGGCGACAGGTATGGCCCTTCCAGGGTAGGA 1532
Db 323 CTTTCATCAGGTTGCTATGGGGCCACGGCGACAGGTATGGCCCTTCCAGGGTAGGA 264
Qy 1533 GGCATTTATCATCCACAGGAAACCCAGAGTATTAAAGAGCCCTGTGGGGCGACAGAC 1592
Db 263 GGCATTTATCATCCACAGGAAACCCAGAGTATTAAAGAGCCCTGTGGGGCGACAGAC 204
Qy 1593 ATAGCAGGGTGGGAGTGCCTCCCTTTATCCTTATCCTGACAAATCTCTAGTCGATCTTTCGCTTT 1652
Db 203 ATAGCAGGGTGGGAGTGCCTCCCTTTATCCTTATCCTGACAAATCTCTAGTCGATCTTTCGCTTT 144
Qy 1653 TTCTCCGATTGGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCCTGCTCAA 1712
Db 143 TTCTCCGATTGGGATTTGGGGCCACCTCTAAGATGCTCTCTCCAGCCCTGCTCAA 84
Qy 1713 CCATCTCCAAATTAGTGCCAAACCCAGGGGCTTGGCAGCTTCCACATCATCATTCATTCCT 1772
Db 83 CCATCTCCAAATTAGTGCCAAACCCAGGGGCTTGGCAGCTTCCACATCATCATTCATTCCT 25
Qy 1773 GCTGCCAAGTGGAAATAACGGCG 1796
Db 24 GCTSCNMYSYSCMGATMMCSR 1
```

RESULT 8

```
LOCUS BX345478/c 993 bp mRNA linear EST 05-MAY-2003
DEFINITION BX345478 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0C024YF06 3-PRIME, mRNA sequence.
ACCESSION BX345478
VERSION BX345478.1 GI:30383001
KEYWORDS EST.
```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 993)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6579.r
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AC024DC03NP2&cluster=6579.r. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0AC024DC03NP2.

FEATURES

Location/Qualifiers
1..993
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC024YF06"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 49.6%; Score 998.6; DB 13; Length 993;
Best Local Similarity 98.2%; Pred. No. 4.4e-205;
Matches 920; Conservative 0; Mismatches 14; Indels 3; Gaps 1;
Qy 854 GTGGCAAGGAGGATCGAAGGCTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCA 913
Db 943 GGTGGCAAGAGATCGAAGGCTTGGGGCAAGGGAGCTGTGAGACCAACCCCTCA 887
Qy 914 AGGGCAGCACCCCAACCTCACCAAGGAGAGAAACAAATACAGACCATCAGCCAC 973
Db 886 AGGGCAGCACCCCAACCTCACCAAGGAGAGAAACAAATACAGACCATCAGCCAC 827
Qy 974 ACCCGCTTACTGTGATGAGTCCGTGTTGGCTCCGATCTGAGGGCGCAGCTTCGGG 1033
Db 826 ACCCGCTTACTGTGATGAGTCCGTGTTGGCTCCGATCTGAGGGCGCAGCTTCGGG 767
Qy 1034 GCCCGCGGATGGGAGAGGGGATGCCCAAGCTCCGTGCTCTCTTGTGACGCCACA 1093
Db 766 GCCCGCGGATGGGAGAGGGGATGCCCAAGCTCCGTGCTCTCTTGTGACGCCACA 707
Qy 1094 CTTACCCCAAGGGTAGCCTTCGCGCCCGCCAGGGAGGACCACTCGAGGCCATTAC 1153
Db 706 CTTACCCCAAGGGTAGCCTTCGCGCCCGCCAGGGAGGACCACTCGAGGCCATTAC 647
Qy 1154 CCAGCTGCTCTCCCAAGACAGCGGGGCGAGGGCAGACTCCAGAGTTATCTATG 1213
Db 646 CCAGCTGCTCTCCCAAGACAGCGGGGCGAGGGCAGACTCCAGAGTTATCTATG 587
Qy 1214 GGTGGTTTACACTCTTTCACGCCCTTGAAGCGGGAGCTTTCCCATTCCTCACCACCTG 1273
Db 586 GGTGGTTTACACTCTTTCACGCCCTTGAAGCGGGAGCTTTCCCATTCCTCACCACCTG 527
Qy 1274 AATGTCCCAAGCTGGTTCATCCAGCCACCTGCGCCCAACAAATGGGCTTCAGAT 1333
Db 526 AATGTCCCAAGCTGGTTCATCCAGCCACCTGCGCCCAACAAATGGGCTTCAGAT 467
Qy 1334 CTCAGGCTTTCACAGTCAAGGGGTGACCTTCCGAGGCCCTCGTGACTTCCAGGGCTCGC 1393
Db 466 CTCAGGCTTTCACAGTCAAGGGGTGACCTTCCGAGGCCCTCGTGACTTCCAGGGCTCGC 407

Qy	1394	TCAGTTAGCAATTTTCAGTGGCCATCTACCCACGACGAGGTGGGCCACCCAGAAACCAAAG	1453
Db	406	TCAGTTAGCAATTTTCAGTGGCCATCTACCCACGACGAGGTGGGCCACCCAGAAACCAAAG	347
Qy	1454	CCCCCTTGGAAATGATATCTCTTTTCATCAGGGTTCCTATGGGCGCCAGCGGACAGGTATG	1513
Db	346	CCCCCTTGGAAATGATATCTCTTTTCATCAGGGTTCCTATGGGCGCCAGCGGACAGGTATG	287
Qy	1514	GCCCTTGCACGGGTAGGAGACATTCATCACCCAGGGAACCCAGGTATTTAAAGAGCC	1573
Db	286	GCCCTTGCACGGGTAGGAGACATTCATCACCCAGGGAACCCAGGTATTTAAAGAGCC	227
Qy	1574	CTGTGGGGGCGACACATAGCAGGGTGGGAGTGCCTCCCTTTATCTGTACATCT	1633
Db	226	CTGTGTGGGGCGACACATAGCAGGGTGGGAGTGCCTCCCTTTATCTGTACATCT	167
Qy	1634	CTAGTCGATCTTTCGCCCTTTTCTCCGATTCGCGATTGGGGGCCACTCTTAAGATGCT	1693
Db	166	CTAGTCGATCTTTCGCCCTTTTCTCCGATTCGCGATTGGGGGCCACTCTTAAGATGCT	107
Qy	1694	CTCTCCAGCCCTGTCTCAACATCTCCAAATAGTGCACCAACCCAGGGGCTTGCACCTC	1753
Db	106	CTCTCCAGCCCTGTCTCAACATCTCCAAATAGTGCACCAACCCAGGGGCTTGCACCTC	47
Qy	1754	CCACATCATCCATTTCTTCTGTCGCCAAGTGCAGTAA	1790
Db	46	CCACATCATCCATTTCTTCTGTCGCCAAGTGCAGTAA	10
RESULT 9	AL582549/c		
LOCUS	AL582549	1030 bp	mRNA linear EST 01-JUN-2003
DEFINITION	AL582549 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED		
ACCESSION	Homo sapiens cDNA clone CS0DL002YI05 3-PRIME, mRNA sequence.		
VERSION	AL582549		
KEYWORDS	AL582549.2 GI:31320766		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	1. (bases 1 to 1030)		
COMMENT	Li, W.B., Gruber, C., Jesses, J. and Polayes, D.		
	Full-length cDNA libraries and normalization		
	Unpublished (2001)		
	On Feb 16, 2001 this sequence version replaced gi:12950640.		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. This sequence belongs to sequence cluster 6579.r For		
	more information about this cluster, see		
	http://www.genoscope.cns.fr/		
	cgi-bin/cluster.cgi?seq=CS0DL002AB03NP1&cluster=6579.r. Contact :		
	Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/ invitrogen Corporation 1600		
	Faraday Avenue Genoscope sequence ID : CS0DL002AB03NP1.		
FEATURES	Location/Qualifiers		
source	1..1030		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DL002YI05"		
	/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"		
	/cell_line="RAMOS CELL LINE"		
	/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT		
	25-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo(dT)		
	primer. Five prime end enriched, double-strand cDNA was		
	digested with NotI and cloned into the Not I and EcoR V		
	sites of the pCMVSPORT 6 vector. Library was normalized."		

[illegible]

```

cgl-bin/cluster.cgi?seq=CS0DL005BH12NP1&cluster=6579.r. Contact :
Feng Liang Email : fliang@lifetech.com/ Invitrogen Corporation 1600
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DL005BH12NP1.

FEATURES
    source
        Location/Qualifiers
            1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DL005Y024"
                /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                /call_line="RAMOS CELL LINE"
                /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                25-NORMALIZED"
                /notes="1st strand cDNA was primed with a NotI-oligo(dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      48.7%; Score 882.2; DB 9; Length 1201;
Best Local Similarity 96.1%; Pred. No. 4.3e-201;
Matches 920; Conservative 19; Mismatches 14; Indels 4; Gaps 4;

Qy 822 CCTGGGTGGAGAGGCTAACAGAACAGAGGCGTGGCAAGGAGGATCGAAGGCTTGG 881
Db 954 CCYTGKTKGAKAGCTTAACAGAACAGAGCGGTGKCAAGGAGGATCGAAGGCTT-G 896
Qy 882 GGGCAAGGGGAGCTGTGAGACCAACCCCTCAAGGGGAGCAGCCCCACCCCTCACACAA 941
Db 895 GGGCAAGGGGAGCTGTGATACCTCCCTCAAGGGGAGCAGCCCCACCCCTCACACAA 836
Qy 942 GGAAGAGAAACAATACAGACCATCAGCCACAC-CCGTCCTTACTGTGATGAGTCGTG 1000
Db 835 GGAAGAGAAACAATACAGACCATCAGCCACACCTCCGTCCTTACTGTGATGAGTCGTG 776
Qy 1001 TTGTGGCTCCGATCTGAAGGCGGACGCTTGGGGGCCCCCGGATGCGAAGGGGATGCC 1060
Db 775 TTGTGGCTCCGATCTGAAGGCGGACGCTTGC-GGGCCCCCGGATGCGAAGGGGATGCC 717
Qy 1061 GCAAGCTCGTCTCTTTGTGGAGCGCACACCTTACCACCCAGGGGTAGCCACTCGGCC 1120
Db 716 GCAAGCTCGTCTCTTTGTGGAGCGCACACCTTACCACCCAGGGGTAGCCACTCGGCC 657
Qy 1121 CGCCCCAGGGAGGACACCTGCGAGCGCATTCACCCAGCTGTCCCTCCAGAGACAGCGG 1180
Db 656 CGCCCCAGGGAGGACACCTGCGAGCGCATTCACCCAGCTGTCCCTCCAGAGACAGCGG 597
Qy 1181 GGGCCAGCGGAGACTCCAGAGAGTTATCTATGGGTGGGTACACTCTTTCAGCCCCCTG 1240
Db 596 GGGCCAGCGGAGACTCCAGAGAGTTATCTATGGGTGGGTACACTCTTTCAGCCCCCTG 537
Qy 1241 AAGCGGGGACTTCCATTCCTCCACCTCAGTGAATGTCCAGAGACTGGTATCCAGCC 1300
Db 536 AAGCGGGGACTTCCATTCCTCCACCTCAGTGAATGTCCAGAGACTGGTATCCAGCC 477
Qy 1301 ACCAGTGCCTCCACCAATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACC 1360
Db 476 ACCAGTGCCTCCACCAATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACC 417
Qy 1361 TTCGGAGGCCCTCGTGTGATCTCCAGGGCTCGCTCAGTTAGCATTTTCACTGCTTACC 1420
Db 416 TTCGGAGGCCCTCGTGTGATCTCCAGGGCTCGCTCAGTTAGCATTTTCACTGCTTACC 357
Qy 1421 CCACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATCTTTTATC 1480
Db 356 CCACGACGAGTGGGGCCACCCAGAAACCAAGCCCCCTTGGAAATGATCTTTTATC 297
Qy 1481 AGGGTTGCCATATGGGGCCACCGGACAGATATGGGCCCTTGCACGGGTAGGAGCATTC 1540
Db 296 -GGGTTGCCATATGGGGCCACCGGACAGATATGGGCCCTTGCACGGGTAGGAGCATTC 238
Qy 1541 ATCACCCAGGGAACCCCGAGTATTAAAGAGCCCTGTGGGGGCGACAGACATAGCAGG 1600

```

```

Db 237 ATCACCCAGGGAACCCAGGCTATTAAAGAGGCCCTGTGGGGGAGACACATAGCAGG 178
Qy 1601 GGTGGGAGTGCCTCCCTTTATCTGAGCAATCTCTAGTCGATCTTCTGCTTTCTCCCG 1660
Db 177 GGTGGGAGTGCCTCCCTTTATCTGAGCAATCTCTAGTCGATCTTCTGCTTTCTCCCG 118
Qy 1661 ATTGGGATTTGGGGGCGCACCTCTAAGATGCCTCTCTCCAGCCCTGTCTCAACCACTC 1720
Db 117 ATTGGGATTTGGGGGCGCACCTCTAAGATGCCTCTCTCCAGCCCTGTCTCAACCACTC 58
Qy 1721 CAATATTAGTCCACCCAGGGGCTGSCACTCCCATCATCATCATCTGCTTCTGCTG 1777
Db 57 CAATATTAGTCCACCCAGGGGCTGCTGCTCTTAACATCATCATCTGCTTCTGCTGCT 1
RESULT 12
AL517881
LOCUS      1201 bp      mRNA      linear      EST 12-MAY-2003
DEFINITION Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CSODA004YD23 5-PRIME, mRNA sequence.
ACCESSION      AL517881
VERSION      AU517881.2 GI:30535630
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1201)
AUTHORS      Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001).
COMMENT      On Feb 13, 2001 this sequence version replaced gi:12781374.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6579.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgl-bin/cluster.cgi?seq=CS0DA004CB12QP1&cluster=6579.r. Contact :
            Feng Liang Email : fliang@lifetech.com/ Invitrogen Corporation 1600
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Paradise Avenue Genoscope sequence ID : CSODA004CB12QP1.

FEATURES
    source
        Location/Qualifiers
            1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSODA004YD23"
                /tissue="NEUROBLASTOMA"
                /clone_lib="Homo sapiens NEUROBLASTOMA"
                /notes="vector: pCMVSPORT 6; 1st strand cDNA was primed
                with a NotI-oligo(dT) primer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                Library was not normalized."

ORIGIN
Query Match      48.6%; Score 880.6; DB 9; Length 1201;
Best Local Similarity 96.3%; Pred. No. 1e-200;
Matches 916; Conservative 14; Mismatches 18; Indels 3; Gaps 3;

Qy 157 TCAGCATGCTGCTGCTGCTCCCTGCTGCTGGGCAAGGCTGCTGCTTCTGGGTTTC 216
Db 64 TCAGCATGCTGCTGCTGCTCCCTGCTGCTGGGCAAGGCTGCTGCTTCTGGGTTTC 123
Qy 217 TGGATTGCGGGGCGGTTTACACGCTAGCTGTGCCGCTCTCTGGGTGAGTCGCTCGCGC 276
Db 124 TGGATTGCGGGGCGGTTTACACGCTAGCTGTGCCGCTCTCTGGGTGAGTCGCTCGCGC 183
Qy 277 GCGGTGCCCCCGGACGCGCTAGGCTCGCGGGGTCCGGGGCCCCCGGCGCTTCGCGGCTGC 336

```

Db	184	GC	GT	G	T	G	C	C	C	G	G	A	G	C	G	C	T	A	G	C	T	G	C	G	G	G	T	C	G	G	G	C	C	C	A	G	C	A	T	T	C	G	G	G	T	G	C		243					
Qy	337	AG	AT	T	C	A	G	C	G	G	G	A	T	C	C	C	G	G	A	T	C	C	C	G	C	G	C	C	T	C	A	C	C	A	G	C	G	G	T	C	C	A	G		394									
Db	244	AG	AT	T	C	A	G	C	G	G	G	A	T	C	C	C	G	G	A	T	C	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		303									
Qy	395	AC	T	G	T	G	G	A	A	G	A	G	T	G	C	G	G	A	C	G	G	T	C	C	T	C	C	G	A	T	C	C	C	C	C	C	C	C	C	C	C	C	C		454									
Db	304	AC	T	G	T	G	G	A	A	G	A	G	T	G	C	G	G	A	C	G	G	T	C	C	C	T	C	C	G	A	T	C	C	C	C	C	C	C	C	C	C	C	C		363									
Qy	455	AT	G	T	C	A	G	C	T	T	A	T	A	G	C	T	G	A	C	T	A	C	A	T	G	A	C	T	T	C	A	C	C	T	C	A	G	T	T	T	G	T	G	A	T	C	C	G	T	A		514		
Db	364	AT	G	T	C	A	G	C	T	T	A	T	A	G	C	T	G	A	C	T	G	A	C	T	T	C	A	C	T	T	C	A	C	C	T	C	A	G	T	T	T	G	T	G	A	T	C	C	G	T	A		423	
Qy	515	AA	T	G	A	C	AA	T	T	C	G	A	A	G	C	T	A	C	A	G	T	C	T	C	T	T	G	A	G	A	T	T	A	A	T	G	A	A	C	A	A	T	G	C	T		574							
Db	424	AA	A	T	G	A	C	AA	T	T	C	G	A	A	G	C	T	A	C	A	G	T	C	T	T	G	A	A	A	G	A	T	T	A	A	T	G	A	A	C	A	A	T	G	C	T		483						
Qy	575	T	G	T	A	A	A	G	C	T	T	T	C	A	G	A	G	G	A	C	C	T	C	G	A	A	G	C	A	G	G	C	C	T	G	G	C	C	G	C	A	G	A	C	A	C	T	G		634				
Db	484	T	G	T	A	A	A	G	C	T	T	T	C	A	G	A	G	A	G	A	C	C	T	C	G	A	A	R	C	A	R	G	C	C	T	G	G	C	C	G	C	A	G	A	C	A	C	T	G		543			
Qy	635	C	T	G	T	C	A	C	A	G	G	A	C	C	A	C	A	G	A	C	A	T	A	A	G	A	C	C	C	C	T	G	A	G	C	T	G	G	C	T	A	G	C	T	A	G	G	A	T	G		694		
Db	544	C	T	G	T	M	A	C	A	C	G	G	A	C	C	A	C	A	G	A	C	C	C	C	T	G	A	A	C	C	C	C	T	G	A	G	C	T	G	G	C	C	T	C	A	G	T	A	G		603			
Qy	695	C	A	G	A	C	C	T	C	G	C	T	T	C	A	G	A	C	C	G	T	C	G	C	A	G	G	T	G	G	C	T	A	C	C	G	G	T	C	A	G	G	C	C	A	G	A	C	T	C	A		754	
Db	604	C	A	G	A	C	C	T	C	G	C	C	T	T	C	A	G	A	C	C	G	T	C	G	C	A	G	G	T	G	G	C	T	A	C	C	G	G	T	C	A	G	G	C	C	A	G	A	C	T	C	A		663
Qy	755	T	A	T	G	T	G	A	T	G	A	G	A	C	T	C	T	G	T	T																																		

RESULT 13

EX384471	EX384471	1201 bp	linear	EST 08-MAY-2003
LOCUS	EX384471	Homo sapiens HELA CELLS	COT 25-NORMALIZED	Homo sapiens
DEFINITION	cDNA clone CS0DK010VL01 5-PRIME, mRNA sequence.			
ACCESSION	EX384471			
VERSION	EX384471.1	GI:30447270		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 1201) Li, W.B., Gruber, C., Jesse, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope			

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 1201)
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 131 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6579.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0C010AG05QPL&cluster=6579.r. Contact :
 Peng Liang Email: fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DC010AG05QPL.
 Location/Qualifiers
 1. 1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DC010YMO9"
 /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
 source
 1. 1201

ORIGIN
 Query Match 47.7%; Score 863.6; DB 13; Length 1201;
 Best Local Similarity 98.3%; Pred. No. 1.3e-196;
 Matches 891; Conservative 2; Mismatches 11; Indels 2; Gaps 2;
 507 GATCCGTAAATGACAAATTCGAAGCTACTTACAGTCTGTTGAGAGGATTAAATGAA 566
 Db 48 GGTCCGGAATTCCTGGGATTTCGAAGCTACTTACAGTCTGTTGAGAGGATTAAATGAA 107
 567 ACAATGCTTCTAAAGCTCTTTGAGGAGGAGGCTCGGAAGCAGGGCCCTGGCGGGCAGAG 626
 Db 108 ACHATGCTTGAAGCTCTTTGAGGAGGAGGCTCGGAAGCAGGGCCCTGGCGGGCAGAG 167
 627 CACACTGTGTCACCAAGGACCAAGGAGCATGAAGACCCCGTGGAGCTGGCCGTCA 686
 Db 168 CACACTGTGTCACCAAGGACCAAGGAGCATGAAGACCCCGTGGAGCTGGCCGTCA 227
 687 GTGGATGACAGCCCTCGGCTTCAGCACCGCTGCCGAGTGGCTACCGGGTCAAGGCCA 746
 Db 228 GTGGATGACAGCCCTCGGCTTCAGCACCGCTGCCGAGTGGCTACCGGGTCAAGGCCA 287
 747 GGACGTCAATATGTGGATGAGACTCTGTTGGCAGCCCGCAGGACCCCGCCCTACCCAC 806
 Db 288 GGACGTCAATATGTGGATGAGACTCTGTTGGCAGCCCGCAGGACCCCGCCCTACCCAC 347
 807 CGGACTTCGATCGCCCTGGTGGAGAGGCTTAACAGACCAGAGGCTGGGCAAGGAGG 866
 Db 348 CGGACTTCGATCGCCCTGGTGGAGAGGCTTAACAGACCAGAGGCTGGGCAAGGAGG 407
 867 CATCGAAGGCTTGGGGGCAAGGGAGCTGTGAGACACCCCTCAAGGGGCGAGCACCC 926
 Db 408 CATCGAAGGCTTGGGGGCAAGGGAGCTGTGAGACACCCCTCAAGGGGCGAGCACCC 467
 927 CCACCTTCACCAAGGAAGAAACAAATACAGACCCCATCAGCCACACCCCGTCTTACT 986
 Db 468 CCACCTTCACCAAGGAAGAAACAAATACAGACCCCATCAGCCACACCCCGTCTTACT 527
 987 GTGATGATCGCTGTTGGCTCCGATCTCGAGCGCCAGCTTCGGGGCCCGCGGATGG 1046
 Db 528 GTGATGATCGCTGTTGGCTCCGATCTCGAGCGCCAGCTTCGGGGCCCGCGGATGG 587

QY 1047 CGAAGGGGATGCGGCAAGCTCCGTGCTCTTGTGGAGCGCCACCACTACCCCAAGG 1106
 Db 588 CGAAGGGGATGCGGCAAGCTCCGTGCTCTTGTGGAGCGCCACCACTACCCCAAGG 647
 QY 1107 GTAGCCACTGCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTACCCAGCTGGTCCCT 1166
 Db 648 GTAGCCACTGCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTACCCAGCTGGTCCCT 707
 QY 1167 CCAAGACAGAGCGCGGGCCAGCGGAGACTCCCAAGAAAGTTATCTATGGTGGGTTACACT 1226
 Db 708 CCAAGACAGAGCGCGGGCCAGCGGAGACTCCCAAGAAAGTTATCTATGGTGGGTTACACT 767
 QY 1227 CTTCAAGCCCTCTGAAAGCGGGGAGCTTCCCAATTCCTTACCCACCTGAAATGTCCTCCAGCA 1286
 Db 768 CTTCAAGCCCTCTGAAAGCGGGGAGCTTCCCAATTCCTTACCCACCTGAAATGTCCTCCAGCA 827
 QY 1287 CTGGTCAATCCAGCCACCAAGTCCCGCCCAACAAATGGGCTTCAGGATCTCAGGGCTTTCCA 1346
 Db 828 CTGGTCAATCCAGCCACCAAGTCCCGCCCAACAAATGGGCTTCAGGATCTCAGGGCTTTCCA 887
 QY 1347 CGTCAGGGGTGACCTTCGGAGCCCTCTGGTGAATTCAGGGCTTCGTCAGTTAGCATTT 1406
 Db 888 CGTCAGGGGTGACCTTCGGAG-CCCCCTGGTGAATTCAGGGCTTCGTCAGTTAGCATTT 945
 QY 1407 CAGTGC 1412
 Db 946 MAGTGC 951

Search completed: August 17, 2004, 09:42:22
 Job time : 4890 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 11:57:41 ; Search time 54 seconds
(without alignments)
1407.505 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTPVELAVSGMOTLGLQHR.....SVETPRRGATQKPKPKWK 269

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1437	100.0	269	4	AAB67330 Human neu
2	1437	100.0	269	4	AAB94540 Human pro
3	863	60.1	177	4	AAB94561 Human pro
4	239	16.6	467	4	ABB11341 Human typ
5	218	15.2	65	5	ABP35240 Human ORF
6	136.5	9.5	4019	4	AAE13839 Human lun
7	136.5	9.5	4019	7	ADD66733 Human lun
8	136.5	9.5	4019	7	ADE87987 Human lun
9	136.5	9.5	4025	5	ABP69736 Human pol
10	136	9.5	2703	4	ABG63299 Drosophil
11	136	9.5	2703	5	ABG70019 Larval vi
12	134.5	9.4	19938	6	ABP76681 Streptomy
13	134	9.3	502	6	ADA24266 Human WAS
14	132.5	9.2	470	4	ABG21932 Novel hum
15	131.5	9.2	705	4	AAM39328 Human pol
16	111.5	9.2	714	4	AAM41114 Human pol
17	129.5	9.0	791	5	ABP5089 Hypoxia-r
18	129.5	9.0	863	6	ABP98856 Human str
19	129.5	9.0	863	7	ADC31071 Human nov
20	129.5	9.0	863	7	ADE48296 Human MIC
21	129.5	9.0	863	7	ADE48308 Human MIC
22	128	8.9	504	2	AAV24091 Human wis
23	128	8.9	566	6	ADA24286 Protein r
24	126.5	8.8	1199	7	ADD46005 Rat Prote
25	126.5	8.8	1199	7	ADE61171 Rat Prote

ALIGNMENTS

RESULT 1

AAB67330

ID AAB67330 standard; protein; 269 AA.

XX

AC AAB67330;

XX

DT 23-APR-2001 (first entry)

XX

DE Human neuron progenitor cell clone #2 protein.

XX

KW Neuron; progenitor cell; gene therapy.

XX

OS Homo sapiens.

XX

FN WO200107607-A2.

XX

PD 01-FEB-2001.

XX

PF 21-JUL-2000; 2000WO-JP004895.

XX

PR 23-JUL-1999; 95JP-00209817.

PR

PR 18-OCT-1999; 99US-0159528P.

XX

(HELI-).HELIX RES INST.

PA

Ota T, Isogai T, Nishikawa T, Kawai Y;

XX

WPI; 2001-182791/18.

DR

New human polynucleotides, particularly DNAs, isolated from a cDNA

PT library derived from progenitor cells, useful in gene therapy, as well as

PT in producing proteins useful as diagnostic markers in drug development.

XX

PS Claim 1; Page 36-37; 54pp; English.

XX

The present invention relates to human proteins isolated from clones from

CC neuron progenitor cells. The proteins and the DNA encoding them may be

CC used in the preparation of treatments for diseases associated with the

XX proteins

SQ Sequence 269 AA;

Query Match 100.0%; Score 1437; DB 4; Length 269;

Best Local Similarity 100.0%; Pred.No. 1.8e-114; Indels 0; Gaps 0;

Matches 269; Conservative 0; Mismatches 0;

QY 1 MKTPVELAVSGMOTLGLQHRCGGYRKARTSYVDITLFGSPAGTRTPPDFDPWVEXA 60

|||||

Db 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60
 QY 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
 Db 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
 QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180
 Db 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180
 QY 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFRSPLV 240
 Db 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFRSPLV 240
 QY 241 TSARSVSISVPSTPRGGATQKPKPWK 269
 Db 241 TSARSVSISVPSTPRGGATQKPKPWK 269

RESULT 2

AAB94540
 ID AAB94540 standard; protein; 269 AA.

XX AAB94540;

XX AC 26-JUN-2001 (first entry)

XX DT Human protein sequence SEQ ID NO:15284.

XX DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX KW Homo sapiens.

XX OS EP1074617-A2.

XX PN 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX PR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 15284; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification, where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 269 AA;

XX Query Match 100.0%; Score 1437; DB 4; Length 269;

XX Best Local Similarity 100.0%; Pred. No. 1.8e-114;

XX Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60

Db 1 MKTPVELAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDFDPPWVEKA 60

QY 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120

Db 61 NRTRGVGKEASKALGAKGSCETTPSRGSTPTLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120

QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180

Db 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSRPREAPLRAIHPAGPSKTEPGPAADS 180

QY 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFRSPLV 240

Db 181 QKLSMGGHLSRPLKRGSLHSLTHLNVSTGHPATSAHTNGPDLPSTSGVTFRSPLV 240

QY 241 TSARSVSISVPSTPRGGATQKPKPWK 269

Db 241 TSARSVSISVPSTPRGGATQKPKPWK 269

RESULT 3

AAB94561

ID AAB94561 standard; protein; 177 AA.

XX AAB94561;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:15335.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX PR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 15335; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX CC Sequence 177 AA;

XX CC Query Match 60.1%; Score 863; DB 4; Length 177;

XX CC Best Local Similarity 99.4%; Pred. No. 11e-65;

XX CC Mismatches 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPELVAVSGMOTGLQHRGCGYRVKARTSVYDTEFLFGSPAGTPTPPDPDPWVEKA 60

DB 1 MKTPELVAVSGMOTGLQHRGCGYRVKARTSVYDTEFLFGSPAGTPTPPDPDPWVEKA 60

QY 61 NRTRGVGKEASKALGAGKSCETTPSRGSTPTLPRKNKYRPTSHTPSYCDSELSGSRSE 120

DB 61 NRTRGVGKEASKALGAGKSCETTPSRGSTPTLPRKNKYRPTSHTPSYCDSELSGSRSE 120

QY 121 GASFGAPRMAKGAADAKLALLWTPTTPPGSGHSPPRPREAPL 161

DB 121 GASFGAPRMAKGAADAKLALLWTPTTPPGSGHSPPRPREAPL 161

RESULT 4

ABB11341

XX AC ABB11341;

XX DT 11-JAN-2002 (first entry)

XX DE Human type II procollagen homologue, SEQ ID NO:1711.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor; hematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic; KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; KW antifungal; vulnery; antiulcer.

XX OS Homo sapiens.

XX WO200157188-A2.

XX

PD 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003800.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00360875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

DR N-PSDB; ABA08585.

XX Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer.

XX Claim 20; Page 167-168; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis; cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention

XX Sequence 467 AA;

QY Query Match 16.6%; Score 239; DB 4; Length 467;

Best Local Similarity 29.4%; Pred. No. 6.7e-12;

Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;

QY 5 VELAVSGMOTGLQHR--CRGYR-----VKARTSYVDTEFLFGSPAGTPTPPDPDPW 56

DB 176 MELCUAAQAQGLPRTIRGWRLTEPEAWRR--HRRPWGORGAVRPPQGAAPP 232

QY 57 VEKANRTRGVGKEASKALGAGKSCETTPSRGSTPTLTPRKNKYRPTSHTPSYC---DES 113

DB 233 SHQGRRT-----NTDPS--ATPRLT-----VMSRCLAPDLK 261

QY 114 LFGSRSEGAFGAPRMAKGAADAKLALLWTPTTPPGSGHSPPRPREAPLRATHPAGESKTE 173

Db	262	APASGPRGHRGNPQSS-----GALLWTPPTPRGSHSPREAPLRAIHGAGPSKSR	314
Qy	174	PGPAADSQKLSMGLHSSRLKRG-----LSHSL-----THLNVPTGHPATS	216
Db	315	AGASGRLEPVIYGVWVTLFTPEAGTFLIPSPTXMSPALVIQPPVPPTQMGLRISGLPRQG	374
Qy	217	APHYNGQDLRSTSGVTP-----RSPVTSRARSVSISVPSTPR-----	257
Db	375	XP-SGAPWXL-PGLAQLAQCHLPHDEVGPPRNOSPLGND---TLSSGLPMGPRRQWPEL	429
Qy	258	---GG---ATQKP-----KPPW	268
Db	430	ARVGGHSSPREQVLKKPLM	449
RESULT 5			
ABP35240	ID	ABP35240 standard; protein; 65 AA.	
XX	AC	ABP35240;	
XX	DT	08-JUL-2002 (first entry)	
XX	XX	Human ORF4213 protein, SEQ ID NO:8426.	
XX	DE	Human; ORF; open reading frame; ORFX; drug screening; diagnosis;	
KW	KW	disease monitoring; cytokine; cell proliferation; cell differentiation;	
KW	KW	immune modulation; haematopoiesis regulation; tissue growth;	
KW	KW	angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;	
KW	KW	thrombolytic; tumour inhibition; bodily characteristic; fertility;	
KW	KW	behaviour; cancer; proliferative disorder; neurological disorder;	
KW	KW	cardiovascular disease; immune system disorder; organ transplantation;	
KW	KW	tissue growth disorder; tissue regeneration disorder; diabetes mellitus;	
KW	KW	hypothyroidism; cholesterol ester storage disease; infection; vulnery;	
KW	KW	vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;	
KW	KW	neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;	
KW	KW	cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;	
KW	KW	dermatological; analgesic; virucide; antibacterial; fungicide.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
XX	PN	WO200190366-A2.	
XX	PD	29-NOV-2001.	
XX	XX		
XX	PA	24-MAY-2001; 2001WO-US017076.	
XX	PR	24-MAY-2000; 2000US-0206690P.	
XX	XX	(CURA-) CURAGEN CORP.	
XX	PA		
XX	PI	Leach MD, Shimmets RA;	
XX	XX		
XX	DR	WPI: 2002-106200/14.	
XX	DR	N-PSDB; ABN79466.	
XX	XX		
PT	PT	Novel human polypeptides and polynucleotides useful for diagnosing,	
PT	PT	preventing and treating cardiovascular disease, neurodegenerative,	
PT	PT	hyperproliferative disorders and disorders related to organ	
PT	PT	transplantation.	
XX	XX		
PS	PS	Claim 10; Page 2350; 2508pp; English.	
XX	XX		
CC	CC	Sequences ABP31028-ABP35561 represent 4534 novel human proteins	
CC	CC	designated ORF (open reading frame) 1-4534, and sequences ABN75054-	
CC	CC	ABN79587 represent cDNAs encoding them. The invention also encompasses	
CC	CC	polypeptides at least 80% identical to the ORF-ORF4534 (collectively	
CC	CC	referred to as ORFX) proteins, polynucleotides at least 85% identical to	
CC	CC	the ORFX nucleic acid sequences, vectors and host cells comprising ORFX	
CC	CC	polynucleotides, the recombinant production of ORFX proteins, antibodies	
CC	CC	specific for ORFX proteins, methods of detecting ORFX polynucleotides and	
CC	CC	polypeptides, methods of screening for modulators of ORFX expression or	

PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer.
XX
PS Disclosure; Page 309-318; 378pp; English.
XX

CC The invention relates to isolated lung tumour-specific proteins and their
CC corresponding cDNA molecules. Lung tumour-specific proteins and their
CC antigen-presenting cells are useful for stimulating and/or expanding T
CC cells specific for a tumour protein, and for inhibiting the development
CC of cancer. The invention also relates to a composition useful for
CC stimulating an immune response, and for treating cancer. The lung tumour
CC specific oligonucleotide is useful in gene therapy and for diagnosis,
CC detection and treatment of lung cancer. The present sequence is human
CC lung tumour-specific protein
XX
SQ Sequence 4019 AA;

Query Match 9.5%; Score 136.5; DB 4; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.053;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
QY 29 ARTSYVDETLFGSPAGTR-----PTPDPFDPWVEK-----ANTRGVGK 68
DB 988 SRPLQWNETTANRPSVRLDSCSSTNNNDPYAKPDTFRVMTDQFPKSLGSLRSPVYSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKVK-----YRPISTHPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFQRIIPDSYARLLTPAPLDGSGPFPKTMQPPPS 1107
QY 109 YCDESLFGSRSEGFASFGAPRMAKGAALKRALLWTPPTPRGSH--SPRPREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRLSVDPIERPAL--TPRPIDNFSHNQSNNDPYSQPLTTPH 1158
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGLHS-----SRPL 194
DB 1159 AVNESFAHPSPRAFSPQGTISRPTSQDPYSQPPGTPRPVVDYSQSSGTARNTDYSQP- 1217
QY 195 KRGSLSHLTHLVNSTGHPATSGAPHTNGPDRLRSTSGVTFERSPLVTSRARSVSISVPST 254
DB 1218 -----PGTPRPTTVDYPSQPPQTPRSTQTDLFVTPVNTQNRHSDPYAHPPGT 1264
QY 255 PRGGATQPKPP 267
DB 1265 PRFGISVPYSQPP 1277

RESULT 7
ADD66733
ID ADD66733 standard; protein; 4019 AA.
XX
AC ADD66733;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human lung tumour-specific related protein, SEQ ID No 425.
XX
KW expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
KW human; lung tumour-specific.
XX
OS Homo sapiens.
XX
FN WO200292001-A2.
XX
PD 21-NOV-2002.
XX
PF 10-MAY-2002; 2002WO-US014975.
XX
XX 11-MAY-2001; 2001US-00854133.
XX
PA (CORI-) CORIXA CORP.
XX
XX Lodes MJ, Wang T, Fan L, Algate PA, Mcneill PD;
PI
XX

DR WPI; 2003-120592/11.

XX New polynucleotide and polypeptide, useful for preparing a composition
PT for diagnosing, treating or preventing cancer.
PT
PS Disclosure; SEQ ID NO 425; 494pp; English.
XX

XX The invention relates to a novel isolated polynucleotide comprising one
CC of 32 47-6080 base pair sequences, given in the specification, or their
CC complements or degenerate variants, at least 20 contiguous residues of a
CC sequence in, or having at least 75 or 90 % identity with the isolated
CC polynucleotide, or that hybridise with the polynucleotide. The invention
CC further comprises: an isolated polypeptide; an expression vector
CC comprising the polynucleotide operably linked to an expression control
CC sequence; a host cell transformed or transfected with the expression
CC vector; an isolated antibody or its antigen-binding fragment that
CC specifically binds to the polypeptide; a method for detecting the
CC presence of a cancer in a patient; a fusion protein comprising the
CC polypeptide; an oligonucleotide that hybridises to the isolated
CC polynucleotide under moderately stringent conditions; a method for
CC stimulating and/or expanding T cells specific for a tumour protein; an
CC isolated T cell population; a composition comprising a first component
CC consisting of carriers and immunostimulants and a second component; a
CC method for stimulating an immune response in a patient; a method for
CC treating cancer in a patient; a method for determining cancer in a
CC patient; a diagnostic kit comprising at least one oligonucleotide or
CC antibody and a detection reagent comprising a reporter group; and a
CC method for inhibiting the development of cancer in a patient. The
CC compositions of the invention have cytostatic activity and can be used to
CC create a vaccine. The isolated polynucleotide is useful for preparing a
CC composition for diagnosing, treating or preventing cancer. This sequence
CC represents a human lung tumour-specific protein relating to the
CC invention.
XX
SQ Sequence 4019 AA;

Query Match 9.5%; Score 136.5; DB 7; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.053;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
QY 29 ARTSYVDETLFGSPAGTR-----PTPDPFDPWVEK-----ANTRGVGK 68
DB 988 SRPLQWNETTANRPSVRLDSCSSTNNNDPYAKPDTFRVMTDQFPKSLGSLRSPVYSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKVK-----YRPISTHPS 108
DB 1048 QTAKGPIAAGTSDHFTKPSRADVFQRIIPDSYARLLTPAPLDGSGPFPKTMQPPPS 1107
QY 109 YCDESLFGSRSEGFASFGAPRMAKGAALKRALLWTPPTPRGSH--SPRPREAPLRAIHP 166
DB 1108 SQDP--YGSVSQ-----ASRLSVDPIERPAL--TPRPIDNFSHNQSNNDPYSQPLTTPH 1158
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGLHS-----SRPL 194
DB 1159 AVNESFAHPSPRAFSPQGTISRPTSQDPYSQPPGTPRPVVDYSQSSGTARNTDYSQP- 1217
QY 195 KRGSLSHLTHLVNSTGHPATSGAPHTNGPDRLRSTSGVTFERSPLVTSRARSVSISVPST 254
DB 1218 -----PGTPRPTTVDYPSQPPQTPRSTQTDLFVTPVNTQNRHSDPYAHPPGT 1264
QY 255 PRGGATQPKPP 267
DB 1265 PRFGISVPYSQPP 1277

RESULT 8
ADE87987
ID ADE87987 standard; protein; 4019 AA.
XX
AC ADE87987;
XX
DT 29-JAN-2004 (first entry)
XX

DE Human lung tumour antigen polypeptide #92.
 XX Human; lung tumour antigen; cancer; lung cancer; CD4+; CD8+; T cell;
 KW immune response; immunostimulant; cytostatic.
 XX Homo sapiens.
 OS US2003118599-A1.
 XX 26-JUN-2003.
 XX 10-MAY-2002; 2002US-00144649.
 XX 02-APR-1999; 99US-00285323.
 PR 09-AUG-1999; 99US-00370838.
 PR 30-DEC-1999; 99US-00476235.
 PR 03-MAR-2000; 2000US-00518809.
 PR 29-MAR-2000; 2000US-00538037.
 PR 05-JUN-2000; 2000US-00588937.
 PR 18-AUG-2000; 2000US-00640878.
 PR 20-SEP-2000; 2000US-00667170.
 PR 01-NOV-2000; 2000US-00704512.
 PR 14-DEC-2000; 2000US-00738973.
 PR 11-MAY-2001; 2001US-00854133.
 XX (CORI-) CORIXA CORP.
 PA Algate PA, Lodes MJ, Wang T, Fan L, Mcneill PD;
 XX WPI; 2003-897103/82.
 DR N-PSDB; ADE87984.
 XX New polynucleotides encode lung tumor antigens and are useful to
 PT stimulate an immune response or detect or treat a cancer in a patient,
 PT particularly lung cancer.
 XX Disclosure; SEQ ID NO 425; 63pp; English.
 XX The invention relates to polynucleotides encoding lung tumour antigens.
 CC The invention also relates to the polypeptides encoded by the
 CC polynucleotides, isolated antibodies or antigen-binding fragments that
 CC specifically bind the polypeptides and a method for detecting cancer in a
 CC patient, comprising obtaining a biological sample from the patient,
 CC contacting the sample with a binding agent that binds a polypeptide of
 CC the invention, detecting in the sample an amount of polypeptide that
 CC binds to the binding agent, and comparing the amount of polypeptide to a
 CC predetermined cut-off value. T cells specific for a tumour protein can be
 CC stimulated and/or expanded by contacting the T cells with a polypeptide,
 CC polynucleotide or an antigen-presenting cell that expresses a
 CC polypeptide. Cancer development can be inhibited by incubating CD4+
 CC and/or CD8+ T cells isolated from a patient with a polypeptide,
 CC polynucleotide or an antigen-presenting cell that expresses a
 CC polypeptide, so that the T cells proliferate. The invention is used to
 CC stimulate an immune response or to detect or treat a cancer in a patient,
 CC particularly lung cancer. This sequence represents a human lung tumour
 CC antigen polypeptide of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification but was obtained in
 CC electronic format from USPTO at seqdata.uspto.gov/sequence.html.
 XX Sequence 4019 AA;
 SQ

Query Match 9.5%; Score 136.5; DB 7; Length 4019;
 Best Local Similarity 24.0%; Pred. No. 0.053;
 Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
 29 ARTSVVDETLFGSPAGTR-----PTPDDPDPKPVK-----ANTRGVGX 68
 988 SFLQNNETNANRPFVRLDLCSSSTNDNPAKPDTPRPVMTQDFPKSLGLSRSPVVS 1047
 69 EASKALGAKGSC-----ETTPSRGSTPTLTPRKXNK-----YRPISHTPS 108
 1048 QTAKGPFAAGTSDHTKTSRADVFQRIQIPDSYARPLTTPAPLDSGPGPKTPMQPPS 1107

QY 103 YCDESLFGSRSEGASFGAPRMAGDAKLRALLWTPPTPRGSH--SPRPREAPLRAHP 166
 DB 1108 SQDP--YGSVSO-----ASRLSVDPYERFAL--TPRPIDFNFSNQNDYDYSQPLTFHP 1158
 QY 167 -----AGPSK--TEPG-----PAADSQKLSMGLHS-----SRPL 194
 DB 1159 AVNESFAHPSRAFSQPGTISRPTSQDPYSQPPGTPRPVWDSYSQSCTARSNTDPYSQP- 1217
 QY 195 KEGLSHSLTHLVNVPSTGHPATSAPTHNGPDLPSTSGVTFRSPLYTSRARSYSISVPST 254
 DB 1218 -----EGTPRPITVDYISQOPQTPRPSTOTDLFTVPTVQRHSDPYAHPPT 1264
 QY 255 PRGGATQKPKPP 267
 DB 1265 PRGISVPSQPP 1277
 RESULT 9
 ABP69736
 ID ABP69736 standard; protein; 4025 AA.
 AC ABP69736;
 XX 20-JAN-2003 (first entry)
 DE Human polypeptide SEQ ID NO 1783.
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacteria;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 XX Homo sapiens.
 XX WO200270539-A2.
 XX 12-SEP-2002.
 XX 05-MAR-2002; 2002WO-US005095.
 XX 05-MAR-2001; 2001US-00799451.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
 XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 XX Wehrman T, Wang J, Wang D, Dmanac RT;
 XX WPI; 2002-759812/82.
 DR N-PSDB; ABZ11953.
 XX New polynucleotides comprising sequences assembled from expressed
 XX sequence tags (ESTs), useful for treating cell-proliferative,
 XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
 XX or coagulation disorders.
 XX Claim 9; SEQ ID NO 1783; 1012pp + Sequence Listing; English.
 XX The invention relates to an isolated polynucleotide (I) comprising a
 XX nucleotide sequence selected from any of 948 sequences (ABZ1119-
 XX ASZ12066) or their mature protein coding portion, active domain coding
 XX protein or complementary sequences. The polynucleotides are useful for
 XX identifying expressed genes or for physical mapping of human genome. The
 XX encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
 XX markers, as a food supplement, for generating antibodies, in medical
 XX imaging, screening and diagnostic assays and for treating cell-
 XX proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 XX or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 XX diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,

CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4025 AA;

Query Match 9.5%; Score 136.5; DB 5; Length 4025;
Best Local Similarity 24.0%; Pred. No. 0.053;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSYVDETLFGSPAGTR-----PTPDDFPVWEK-----ANRTRGVCK 68
Db 994 SRPLQNETTANRPSVRDLCSSTTNDPYAKPDPTRPVMTDQPKSLGSRSPVSE 1053
QY 69 EASKALGAGKSC-----ETTPSRGSTPTLTTPRKNK-----YRPISHTPS 108
Db 1054 QTAGPIAAGTSHFTKPSRADVFORQIPDISYARPLTAPLDSGPGFPTKQPPPS 1113
QY 109 YCDESLFGSRSEGAFCAPRMAKDAKRALIWLTPPTTGRSH--SPRPREAPLRAIHP 166
Db 1114 SQPP--YGSVQ-----AGRLSVDPYERPAL--TPRPIDNFSHQSNNDPYQPLTTPH 1164
QY 167 -----AGPSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194
Db 1165 AVNESFAHPSRAFSQPGT-SRPTSQDPYQPGTGPTRPVVDSYQSSGTRARNTDPYSQ- 1223
QY 195 KRGLSHSLTLNVPSTGHATSPATHTNGPDLPSPSTGVTFRSPVTSRARSVSISVPST 254
Db 1224 -----PGTRPTTVDPYQQQTPRPSTQTLFVTPVNRHSDEYAHPPGT 1270
QY 255 PRGGATQPKPP 267
Db 1271 PRPGISVPYQPP 1283

RESULT 10

ID ABB63299 standard; protein; 2703 AA.
XX
AC ABB63299;
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16689.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07402.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 16689; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU16176-ABU30511) and the encoded proteins (ABU57737-
CC ABU72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2703 AA;

Query Match 9.5%; Score 136; DB 4; Length 2703;
Best Local Similarity 22.5%; Pred. No. 0.036;
Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYVDETLFGSPAGTRPTPDF-----DPPWVEKANRTRGVG 67
Db 633 GGYKMGPGGPGAGQYPPQPOQYPPGNYPRRPQYPPGAYATGPPPTTQA-----GAG 688
QY 68 KEASKALGAKGSCETTPSRGSTPTLTTPRKNKYRPISTHPSYCYDESLFGSRSEGAFCAP 127
Db 689 GANSMPGSAQAG--GYPRG-----MENHTGQYPPYQWVPPSPQOTVPGGAPGAMVGNH 741
QY 128 RVAKGDAKRALIWLTP-----PPTPRGSHSPRPREAPLRAIHPAGPSKTEFGPAADS 180
Db 742 VQGRG-----TPPPVVGPPPPQSGSPRLNLYLKQHLQHKGGVGGSTPPQGP 791
QY 181 QKLSNG--GLHSSRPL-----KRGLSHSLTLNVPST-----GHPATSAFHT 220
Db 792 QGYNGPTGMPPGMPPHMGPPHGTNMGPTSTPPQSQMLQGGQPOQSGASGGPES 851
QY 221 NGPDLPSTSGVTFRSPVTSRARSVSISVPSTP 255
Db 852 GGPEHI-SQDNGISSSGPTGAAGMHAVTSVVTGTP 885

RESULT 11

ID ABB70019 standard; protein; 2703 AA.
XX
AC ABB70019;
DT 05-NOV-2002 (first entry)
XX
DE Larval viability associated protein #18.
XX
KW Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;
KW oilseed rape; soybean; vegetable crop; fruit.
XX
OS Drosophila melanogaster.
XX
PN WO200257455-A2.
XX
PD 25-JUL-2002.
XX
PF 18-JAN-2002; 2002WO-US001568.
XX
PR 18-JAN-2001; 2001US-0262351P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Stam L, Bachmann J, Broadus J, Kamdar KP;
XX
DR WPI; 2002-590746/63.
DR N-PSDB; ABB51395.
XX
XX Identifying inhibitors of activity of proteins essential for Drosophila
PT larval viability comprises expressing in a host a protein essential for
PT larval activity and identifying compounds that inhibit or interact with

the protein.

XX Claim 1; Page 108-117; 169pp; English.

XX The invention describes a method of identifying compounds that inhibit the activity of, or that interact with a protein essential for *Drosophila* larval viability comprising expressing in a recombinant host a DNA molecule to produce a protein essential for larval viability. The method is useful for identifying compounds with insecticidal activity. Compounds identified are useful as insecticides in crops such as maize, wheat, oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar beet, oilseed rape, soybeans, vegetable crops and fruits. This is the amino acid sequence of a fruit fly larval viability associated protein

XX Sequence 2703 AA;

Query Match 9.5%; Score 136; DB 5; Length 2703;
Best Local Similarity 22.5%; Pred. No. 0.036;
Matches 62; Conservative 30; Mismatches 119; Indels 64; Gaps 11;

QY 23 GGYRV-----KARTSYVDLFLGSPAGTPTPDF-----DPPNVEKANTRGVG 67
DB 633 GGYRVGGGSGAQGPPQPOQYPPGNNPPRPPQPPGAYATGPPPTTQA-----GAG 688
QY 68 KEASKALGAKGSCETTPSRGSTTLTPRKKNKYRPISTPSYCDLSFGSRSEGASFGAP 127
DB 689 GANSMPSGAQAG-GYVGRG-----MENHTGQYPPYQWVPPSPQQTVPGAGPGAMVGNH 741
QY 128 RMAGDAKALRALLWTP-----PPTPGSHSPREAPLRAIHPAGSKTEPGPAADS 180
DB 742 VQKGK-----TPPPFVVGPPPPQSGSPRELNYLKQHLQHKGYGSGSPPTPQGP 791
QY 181 QKLSMG--GLHSKRPL-----KRGLSHSLTLNVPST-----GHPATSAPIH 220
DB 792 QGYGNGPTGHPGMPGMPGHPHMGPPHGTNMGPTTTPPQSQMLQGGQGGASGGPES 851
QY 221 NGQDLRPTSGVTFRSPLVTSRARSVSISVPSTP 255
DB 852 GGPEHI-SQDNGITSSSGPTGAAGMHAIVTSVITGP 885

RESULT 12
ABP76681

XX ID : ABP76681 standard; protein; 19938 AA.
XX AC ABP76681;
XX DT 26-FEB-2003 (first entry)
XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.
XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX OS Streptomyces viridochromogenes.
XX PN WO200268436-A1.
XX PD 06-SEP-2002.
XX PF 24-AUG-2001; 2001WO-BP009815.
XX PR 25-FEB-2001; 2001DB-01009166.
XX PA (COMB-) COMBINATURE BIOPHARM AG.
XX PI Weithauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX WPI: 2003-018650/01.
XX DR N-PSDB; AB237516.
XX PT New avilamycin derivatives, useful for treatment of infections, and nucleic acid encoding avilamycin synthesis enzymes.

XX Example 1; Page 68-301; 319pp; German.

XX The invention relates to avilamycin derivatives (I) with antibacterial, virucide, protozoacide and fungicide activity. (I) are useful for treatment of infections (bacterial, viral, protozoal or fungal), in human or veterinary medicine, particularly where caused by *Staphylococcus aureus*. (I) are more hydrophilic than known avilamycins. The present sequence is that of an avilamycin synthesis enzyme from the *Streptomyces viridochromogenes* Avilamycin A biosynthetic gene cluster (AB237515-AB237516)

XX Sequence 19938 AA;

Query Match 9.4%; Score 134.5; DB 6; Length 19938;
Best Local Similarity 26.8%; Pred. No. 0.55;
Matches 75; Conservative 22; Mismatches 94; Indels 89; Gaps 15;

QY 49 PPDPDPWVEKANTRGVGKEASKALGAKGSC-----ETTPSR 86
DB 11439 PSELPOQWVE---RXROSETPEPMSASSQRCRTWXRSPAQSVASKWPSPANSDTSASR 11495
QY 87 GSTPTLTP-----RKKNKYRP-----ISHTPSYCDLSFGSRSEGASFGAPRMAKGA 134
DB 11496 SST-LVTPTCTSMNTSARRRFRSPSRARIS-TPG-CPDIARVSAPSSAPSSRTLAPGNA 11552
QY 135 AK-LRALLW-----TPPTPRGSH--SPRPREAPLRAIHPAGSKTEPG 175
DB 11553 AKNASASWGWLSRXSRVEAVRYRTPLYTKSFTHGSSSTRRRISRSFTVALP----PL 11608
QY 176 PAADSQKLSGGGLHSSRPLXRGLSHSLTLNVPSTGHPATSAP-----HTNGPQD- 225
DB 11609 SGALPSETAWAGDHRSAPLRSATH-----CAHSGTRSPPISTKEERSAGPTEA 11659
QY 226 -----LRPSTGVTFRSPLVTSRARSVSISVPSTPRRG 259
DB 11660 AHAARASPSASGVSAQSP-KTSRASSVAVXATITTSRAG 11698

RESULT 13
ADA24266

XX ID : ADA24266 standard; protein; 502 AA.
XX AC ADA24266;
XX DT 20-NOV-2003 (first entry)
XX DE Human WASP interacting protein (WIP) amino acid sequence.
XX KW WASP interacting protein; WIP; transgenic; Wiskott-Aldrich syndrome; WAS;
XX KW X-linked immunodeficiency; WAS protein; WASP; thrombocytopenia; eczema;
XX KW impaired immunity; lymphoma; leukaemia; WIP agonist; WIP antagonist;
XX KW cytosstatic; T cell receptor; T cell activation; B cell proliferation;
XX KW human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX Region 8..13
XX FT /label= APPPPP_motif
XX FT Misc-difference 231..232
XX FT /note= "Encoded by GGAGGAGGC"
XX FT Misc-difference 359..360
XX FT /note= "Encoded by CCAGTGCCC"
XX FT Region 426..431
XX FT /label= APPPPP_motif
XX PN WO2003070893-A2.
XX PD 28-AUG-2003.
XX PR 18-FEB-2003; 2003WO-US004646.
XX PT

```

PR 19-FEB-2002; 2002US-00078547.
XX (CHIL-) CHILDRENS MEDICAL CENT.
PA Ramesh N, De La Fuente MA, Anton IM, Geha RS;
XX
PI WPI; 2003-712614/67.
XX DR N-PSDB; ADA24265.
XX
XX New transgenic non-human mammal, useful for enhancing or inhibiting T
PT cell receptor-mediated T cell activation or B cell proliferation in a
PT mammal for treating or preventing diseases, e.g., leukemia.
XX
XX Example 1; Fig 1D; 106pp; English.
PS
XX This invention relates to a novel WASP interacting protein (WIP) and the
CC DNA sequence which encodes it. In addition, the invention relates to a
CC novel transgenic non-human mammal whose genome comprises a disruption of
CC a WIP gene. Wiskott-Aldrich syndrome (WAS) is an X-linked
CC immunodeficiency caused by mutations that affect the WAS protein (WASP).
CC It is characterised by thrombocytopenia, eczema, impaired immunity and a
CC predisposition to develop lymphomas and leukaemias. Modulation of the
CC activity of a WIP protein, through use of a WIP agonist or WIP
CC antagonist, may produce cytostatic activity. The transgenic non-human
CC mammal may be useful for enhancing or inhibiting T cell receptor-mediated
CC T cell activation or B cell proliferation in a mammal for treating or
CC preventing diseases such as leukaemia. The present sequence is the amino
CC acid sequence of the human WIP protein of the invention.
XX
XX Sequence 502 AA;
XX
XX Query Match 9.3%; Score 134; DB 6; Length 502;
XX Best Local Similarity 25.3%; Pred. No. 0.0067;
XX Matches 66; Conservative 22; Mismatches 89; Indels 82; Gaps 13;
XX
XX 42 PAGTR-----PTPPDFPPVWEKANTRGVGKEASKALGAKGCETTPRGSTP----- 90
XX 207 PGGPRQSPGTTTPPF-----PGRNLTALG-----GSIROSPLSSSPFNRRPP 250
XX
XX 91 -----TLTPRKKNKVRTPSHTPSYCDESLFGSRSEGASFGAPRMAGDAKLRALLWT 143
XX 251 LPTFPRALDCKPPPPPPVGNRPSTHREAV----- 281
XX
XX 144 PPPTPRGSHSPRPR---EAPLRA-IHPAGPSKTEPGPAADSQKLSMGLHSSRLPKRGL 198
XX 282 PPPPPQNNKVPSTPRPSAPHRPHLEPPPPSRPGPPPLPPS---SSGNDETRLPQRL 338
XX
XX 199 SHSLTHLNVSTGH--PATSAPHTNGQDLR--PSTSGVTFRSPLVTSRARSVISVPSST 254
XX 339 SLSSSTPLPSGRSGPLPPPPGSRPPPPVDRPPGRSG-PLPPPPPVSRNGSTSRALPAT 397
XX
XX 255 PR---RGATQK---PKPP 267
XX 398 PQLPSRSGVDSRSGPRPP 416
XX
XX RESULT 14
XX ID ABG21932 standard; protein; 470 AA.
XX AC ABG21932;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21923.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX EN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS86119.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 52291; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 470 AA;
XX
XX Query Match 9.2%; Score 132.5; DB 4; Length 470;
XX Best Local Similarity 25.1%; Pred. No. 0.0083;
XX Matches 66; Conservative 18; Mismatches 94; Indels 85; Gaps 12;
XX
XX 42 PAGTRTPPPDFPPVWEKANTRGVGKEASKALGAKGCETTPRGSTPTLTPRKKNKXR 101
XX 240 PPSATPSPP-----TGSPATGRTSFRTELATPRR-----R 269
XX
XX 102 PISHTPSYCDESLFGSRSEGASFGAPRMAGDAKLRALLWTPPP-----TPRGSH 152
XX 270 AASQADTPCQP---GTPSRARP--SPR-TRPPAARFAPXAWPRPPSVAGARWWTXRGAP 323
XX
XX 153 SP-----RPREAFELRAIHPAG---PSKTEPGPAADSQKLSMGLHSSRLPKRGLSHS 201
XX 324 APPGPPGGRPARLPRAAGPAGRRHRPDRGPPSPAAQRPRAAPRTSRSASR----- 378
XX
XX 202 LTHLNVSTGHPATSA-----PHT-----NGPQDLRSTSGVTFRSPLVTSRARS 246
XX 379 -----APRSAGFAAAAGTATAARPPATQARAPRRAPRRPRTATSSXTARAPPPSGAKS 433
XX
XX 247 VSISVPSTPRRGATQKPKPPWK 269
XX 434 AE-----PPEQPPPPPPPPPKK 452
XX
XX RESULT 15
XX ID AAM39328
XX AAM39328 standard; protein; 705 AA.

```

XX AAM39328;
 XX 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 2473.
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 XX 26-DEC-2000; 2000WC-US034263.
 XX 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-0052317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58484.
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX Example 4; SEQ ID NO 2473; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX Sequence 705 AA;
 SQ
 Query Match 9.2%; Score 131.5; DB 4; Length 705;
 Best Local Similarity 25.2%; Pred. No. 0.017;
 Matches 61; Conservative 18; Mismatches 110; Indels 53; Gaps 8;
 QY 40 GSPAGTRPTPPDPPWVEKANTRG-----VGKEASKALGAKGSCETTPSRGSGTP 90
 DB 454 GPPAQQRPPGPGPPQPGFQRPPLQPPPPQGGQHLGSLGPPAG-SPLPQRLPSP 511
 QY 91 TLTPRKKNKYPISHTPSYCDSESLFGRSEGFAGPRMAKGDAAKRLALLMTPTPTPR- 149

Db 512 TSAPO-----QFASQAAPPQTQGRQSRPVAGFGFAPPAARPPASPPQAGPPQATRQ 566
 QY 150 -----GSHSPRPRAIHPAGPSKTEPGPAADSQKLSMGLHSSRPLKRLGSLHL 205
 DB 567 TSVSGPAPPKASGAPPGGQQQPGPPQKPPGAGPTRQASQAG----- 608
 QY 206 NVPSTGHPTASAPHTNGQDLRPSTSGVTFRSPLVTSRARSVSISVSPTRRGATQPK 265
 DB 609 PVPRTGPPPTQQP-----RPSGPGPAGRPKPQLAQKPSQDVPPPTATAAAG----- 654
 QY 266 PP 267
 DB 655 PP 656
 Search completed: August 10, 2004, 12:11:12
 Job time : 56 secs

Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 28918
 LENGTH: 920
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-28918

Query Match 9.4%; Score 135.5; DB 4; Length 920;
 Best Local Similarity 26.3%; Pred. No. 0.0061;
 Matches 78; Conservative 24; Mismatches 110; Indels 85; Gaps 16;
 QY 5 VELAVSGMOTLGLQHCRCGGRVYKARTSYVDLTLF---GSPAGTRTPDPDFPPWYERAN 61
 DB 651 VLRLVRLQQQFGVVRDRLVGLAIAAGHEY---LFAVEGQFQHR--HPESD----- 696
 QY 62 RTRGVGKEASKALGAKGSCETTPTSGSTPTLTTRKKNKYRPISTHTSYCDESLFGSR--- 118
 DB 697 RQCGFGLQAR-LQMLVSGRRVCHRAPAPATSTGRCRPPGRP--CRPRSAGNRPPR 753
 QY 119 -SEGASGAPRMA-----KGDAKRLALLWTPPTPR-----GSHSRP----- 156
 DB 754 APPCAMAGARRRTRGHRGSGASAGAAGW-PPAPAADRGCFAGPGASHGSHVHSAFG 812
 QY 157 --REAPLRAHPAGSKTEPGPAADSKLSMGLHSSRLKGLSHSLTHLNVSTGHPA 214
 DB 813 AGARPLPWSGEPANPAGGA-----PCARRSR-----A 845
 QY 215 TSAPHTN---GPDLRPTSGVTFRPLVTSRARSVSISVPSTPRRGGATQPKPDPW 268
 DB 846 TASCHRHAWGWRDAGLPGAGCS-RESTATTTDA-----PPASPRSGAWCGSRPAW 895

RESULT 3

US-09-599-287A-2
 Sequence 2, Application US/09599287A
 Patent No. 6635446
 GENERAL INFORMATION:
 APPLICANT: Narayanaswamy Ramesh
 APPLICANT: Ines M. Hartwig
 APPLICANT: John H. Hartwig
 APPLICANT: Raif S. Geha
 TITLE OF INVENTION: WIP, A WASP-Associated Protein
 FILE REFERENCE: 1242.1022-004
 CURRENT APPLICATION NUMBER: US/09/599,287A
 CURRENT FILING DATE: 2000-06-22
 PRIOR APPLICATION NUMBER: PCT/US98/27501
 PRIOR FILING DATE: 1998-12-22
 PRIOR APPLICATION NUMBER: 60/101,457
 PRIOR FILING DATE: 1998-09-23
 PRIOR APPLICATION NUMBER: 60/068,533
 PRIOR FILING DATE: 1997-12-23
 NUMBER OF SEQ ID NOS: 26
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 503
 TYPE: PRT
 ORGANISM: Human
 US-09-599-287A-2

Query Match 9.4%; Score 134.5; DB 4; Length 503;
 Best Local Similarity 25.8%; Pred. No. 0.00035;

Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;
 QY 42 PAGTR-PTPPDPDFPPWVEKANRTRGVGKEASKALGAKGSCETTPTSGSTPT--LTPRKN 98
 DB 207 FGGPRQPSGPTTPTTFCNRTALGGGSIROSLSSSPFSNRPLPTTSRALLDDKPP 266
 QY 99 KYRPISTHTPSYCDLSLFGSRSEGASFGAPRMVAKDAKRLALLWTPPTPRGSHSRPR- 157
 DB 267 PPPPVGNRPSIHRFAV-----PPPPQNNKPPVPT 297
 QY 158 ---EAPLRA-IHPAGPSKTEPGPAADSKLSMGLHSSRLKGLSHSLTHLNVSTGH- 212
 DB 298 PRPSAPHRPHLRPPPPSRGPPPLPPS---SSGNDETFRLPQRLNLSLSSSTPPLPSGRS 354
 QY 213 -PATSAPHTNGQDLR--PSTSGVTFRPLVTSRARSVSISVPSTPR---RGGATQK--- 263
 DB 355 GPLPPPPSERPDPFVRDPPGRSG-PLPPPPPVSRNGSTSRALPATPOLPSRSGVDSRSG 413
 QY 264 PKPP 267
 DB 414 PRPP 417
 RESULT 4
 US-09-252-991A-29419
 Sequence 29419, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 29419
 LENGTH: 1008
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-29419

Query Match 9.2%; Score 132.5; DB 4; Length 1008;
 Best Local Similarity 24.8%; Pred. No. 0.0013;
 Matches 76; Conservative 25; Mismatches 141; Indels 65; Gaps 13;
 QY 17 LQHCRCGGYR---VKARTSYVDLTLFGSP-AGTRPTTDPDFPPWVEKANRTRGVGKEASK 72
 DB 377 LVRRGGRRRPPVVRAGTGHGRRRPPAGQPAAGADRT----DPPQSGPASSARPPARRR 432
 QY 73 ALGAKGSCETTPTSGRSTPTLTTRPKK---NKYRPISTHTPSYCDLSLFGSRSEGASFGAPRM 129
 DB 433 ATGAAGTQPAARGA--ALRPPGAGAGAFRLPGRPTAAQAAPGRTGCGAGSRREV 490
 QY 130 AKGDAKRLALLWTPPTT--PRGSHSRPREAPLRAIHPAGPSKTEPGPAADSKLSMGG 187
 DB 491 AAGLGRRRQPARLRPPPTQFPSSGDPFAARPARRTAALPARPELDADPARTQRRRAGR 550
 QY 188 LHSS-----RPLKGLSHSLTHLNVSTGH-----PATSAPHTN 221
 DB 551 RHGAGQDLAVAGRAPCKAG--RAARHAGAGGDAHQDPQLAGRGRALRPRSRAGARR 608
 QY 222 GPQD-LRP---STSGVTFRPLVTSRARSVSISV-PSTPRRG-----G 259
 DB 609 RPAPRLRPHRRARPAGADHLCPTATRRRRAQAQAVPPADPRRGAERHQHQRHRRRPRAG 668
 QY 260 ATQKPK 266
 DB 669 GTASVP 675

```

Query Match      8.9%; Score 128.5; DB 4; Length 501;
Best Local Similarity 24.3%; Pred. No. 0.0012;
Matches 67; Conservative 25; Mismatches 113; Indels 71; Gaps 12;

QY  2  KTVPELVASGMQTGLQHQHRCRGYRVKATSYVDETLFGSPAG-----TRPTPPDF 52
DB  274  RTPV-----RQGRGTARPARP-----LAAAGPALPARQRPGRPGI 314

QY  53  DPFVWEKANTRGVGKEASKALGAKGSCETTPSRGSTPLTPRKKNKYRPISHTPSYCDE 112
DB  315  -----PGNPQRPVGK---RATGQRQVRPFGPARCAGQAVR-RGRGQDRPAPR----- 357

QY  113  SLFGSRSEGASFGAPMAKGDAAKLALLWTPTTPRGSHSPRPREAPLRAIHPAGFSKT 172
DB  358  ---GTRGQQAQVRRRHGRTGA-----PDRAQGGQPEPRQ---RAGQPAHPARG 402

QY  173  EPGPAADSKQLSGGJHSSRPL-----KR---GLSHSLTHLNVSTGHPATSAPTHNG 222
DB  403  RPEPARQGLPAPRGDPHPGKRLTPGNGCAKREAMGIPHGFHF-ISSSEGHNGWPRGDRS 461

QY  223  PQDLRSTSGVTFERSPLVTSRARSVISVPSTPRRG 258
DB  462  FRSPAPAGPARRRSGIATIALDLRQRSAGREG 497

RESULT 7
US-09-252-991A-16754
; Sequence 16754, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16754
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16754

Query Match      8.9%; Score 128.5; DB 4; Length 536;
Best Local Similarity 21.9%; Pred. No. 0.0013;
Matches 87; Conservative 27; Mismatches 113; Indels 171; Gaps 19;

QY  18  QHRC---RGGYRVKATSYVDETLFGSPAGTRP-----TPPDDFPVWEKANTRG 65
DB  81  RHRCIPASATVRSCTRTSAAPSP---SCAGSTPRNWRRAAAATPFCAGTDQASRTSG 137

QY  66  ---VCKEKASKALGAKGSCETTPSRGSTPLTPRKKNK---VRPI-----SHTPSYC 110
DB  138  ASGQYVSRNASR---QADSCAASGRRRRAKSLQPCRRNRGIGTRPLMASOKASQLHPYH 194

QY  111  DESLFGSR-----QADSCAASGRRRRAKSLQPCRRNRGIGTRPLMASOKASQLHPYH 194
DB  195  GGNARASRRSKCKEALSTSSAMRLAARFSSSSMEPPPPAMRRNARTRTRARGSPGGST 254

QY  125  GAFPMKGDAAKLALLWTPTTPRGSHSPRPREAPLRAIHPA----- 167
DB  255  GRGGYTRGSGASPRPA--TPAAAGAGARRARPRPARTCPRCRPGPSAAPRRRRR 312

QY  168  -----GPSKTEFG-----PAADSKQLSGGLSHSSRPLKRG 198
DB  313  DRRRPSGCGGPAGTSGRVPGRRVGRAGNCARAGRAPAGAARRPG-----RPVRR-- 365

QY  199  SHSLTHLNVSTGHPATSA-----HTNGQDILRPSTSGVTFERSPLVTSRAR---SVSISVPS 253

```

Db 366 -RALAGCPVPAAGRCGWSRPARSGSP---RPPVAG---RDSPGRRAAGAGSVADSAPG 418
QY 254 T-----PRGG-----ATQKPKP 266
Db 419 TRRARAAGCGRSRRAKAAVPRRSRCPAATGAPPP 456

RESULT 8

US-09-599-287A-24
; Sequence 24, Application US/09599287A
; Patent No. 6635446
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Ines M. Anton
; APPLICANT: John H. Hartwig
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-004
; CURRENT APPLICATION NUMBER: US/09/599,287A
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translated WIP ORF No. 6635446 3
US-09-599-287A-24

Query Match 8.9%; Score 128; DB 4; Length 507;
Best Local Similarity 25.3%; Pred. No. 0.0014;
Matches 62; Conservative 23; Mismatches 108; Indels 52; Gaps 11;

QY 42 PACTR-PTPDPDPVWEKANTRGVGKEASKALGAKGSCETTPRSGSTPT--LTPRKKN 98
Db 210 PGGPRQPSGPTPTPPGNGRTALGGSIRQSLSSSSPSFNRPLPPTPSALDDKPPP 269
QY 99 KYRPISTHSYCDSELFGRSESGAFGRMAKGDAAKURALLWTPPTTPRSGSHSPRR- 157
Db 270 PPPVGNRPSIHREAV-----PPPPQNNKPPVPST 300
QY 158 ---EAPLRA-IHPAGSKTEPGPAADSQXLSMGLHSSRLPKRGLSHSLTLNVPSTGHP 213
Db 301 PRSAHRHLRPPPSRPGPPLPS---SSGNETPLPORNLSLSSTPLPSGRS 357
QY 214 A-----TSAPHTNGQDUR--PSTSGVTFRSLVTSRARSVISVPSTPR---RGGATQK- 263
Db 358 GPLPPVPVSPRPPPPVDRDPGRSG-PLPPPPVPSRNGSTSRALPATPOLPSRSGVDSPRS 416
QY 264 -PKPP 267
Db 417 GPRPP 421

RESULT 9

US-09-252-991A-26071
; Sequence 26071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26071
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26071

Query Match 8.8%; Score 126.5; DB 4; Length 693;
Best Local Similarity 26.9%; Pred. No. 0.0028;
Matches 87; Conservative 28; Mismatches 112; Indels 97; Gaps 19;

QY 16 GLQHRCR-----GGVYVKARTSYVDETLFGSP---AGTRPTPPDFDPWVEKANRTR 64
Db 332 GLPQRAQPAATACQPGTAGRQRSGVADP--GGPLOGAGRRLEP-----NVRP 378
QY 65 GVGKEASKALGAKGSCETTPSRGSTPTLTP-----RKKNYRPISTHSYCDSELFGRSE 120
Db 379 GVGLRRRRARTGEGVNGGPAAR--YPLRPRGGARQLHPCRPAGDLS--GDRFAPGRRAR 435
QY 121 CASFGAP-----RVAKGD---AAKLRLWTP-----PPTPRGSH 152
Db 436 TGALGAPAAATQYAAQRGDKRSALLRALQADPTGRRRRRGRRRQVPVPRQRSPAPPRGDR 495
QY 153 S-PPREAP-----LRAIHPAGSKTEPGPA-----ADSOKLSMGLHSSRPL 194
Db 496 TGPRAPDAATGGLPGRPSGRLRRPH-AGPG--HFGPARGPRGRADHPQNLAGLDPCRPR 552
QY 195 KR-GLSHSLTLNVPSTGHPATSAPTHNGP-----ODLRPSTSGVTFRSPLVTSRA 244
Db 553 SRPGVQRGLCRAGIPGTAWRAT---HAGGPAASALPERGRAVLPAFLGLRLRRTRNAARS 609
QY 245 RVSISVPS---TPREGGATOKP 264
Db 610 RRGAGQPSRGLRRRRGRARRLP 633

RESULT 10

US-08-906-865-4
; Sequence 4, Application US/08906865
; Patent No. 6040168
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kac, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,865
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "Synapsin Ia"
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-906-865-4

Query Match 8.8%; Score 126; DB 3; Length 696;
Best Local Similarity 24.1%; Pred. No. 0.0031;
Matches 60; Conservative 21; Mismatches 114; Indels 54; Gaps 10;
QY 40 GSPAGTRTPPDPPWVEKANTRG-----VGKEASKALGAKGSCETTPSRGSTP 90
DB 454 GPPAQQEP-PPQGPPPGPGPQPGPQQPQQPQQHLSGLGPPAG-SPLPQRLPSP 511
QY 91 TLTPRKKNKYRPISTHTSYCDSELFGRSEGFAPRMAGDAKRLALLTTPPTPR- 149
DB 512 TSAFQ-----QPASQAAPPTQGGQSRPVAGGFGAPPAARPPASPSQORAGPPQATRQ 566
QY 150 -----GSHSPRPREAPLRAIHPAGPSKTEPCGPAADSKLSMGLHSSRPLKRGLSHSLTHL 205
DB 567 TSVSGPAPPKASGAPGGQQRQPPQKPGPAGTTRQASQAG----- 608
QY 206 NVPTSGHPATSAHTNGPQDLRSTSGVTFRSPLVTSRARSVISVPTPRRGATQ---- 262
DB 609 FVRTGPTPTTQQRPSG-----PGPAGAP--KPQLAQKP--SQDVPPPPATAAAGGPPHPQL 660
QY 263 ----KPKPP 267
DB 661 FNLPEPAPP 669

RESULT 11
US-09-129-668-4
; Sequence 4, Application US/09129668B
; Patent No. 6429010
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES
; FILE REFERENCE: 600-1-202 CIP
; CURRENT APPLICATION NUMBER: US/09/129,668B
; CURRENT FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 08/906,865
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-129-668-4

Query Match 8.8%; Score 126; DB 4; Length 696;
Best Local Similarity 24.1%; Pred. No. 0.0031;
Matches 60; Conservative 21; Mismatches 114; Indels 54; Gaps 10;
QY 40 GSPAGTRTPPDPPWVEKANTRG-----VGKEASKALGAKGSCETTPSRGSTP 90
DB 454 GPPAQQEP-PPQGPPPGPGPQPGPQQPQQHLSGLGPPAG-SPLPQRLPSP 511
QY 91 TLTPRKKNKYRPISTHTSYCDSELFGRSEGFAPRMAGDAKRLALLTTPPTPR- 149
DB 512 TSAFQ-----QPASQAAPPTQGGQSRPVAGGFGAPPAARPPASPSQORAGPPQATRQ 566

QY 150 -----GSHSPRPREAPLRAIHPAGPSKTEPCGPAADSKLSMGLHSSRPLKRGLSHSLTHL 205
DB 567 TSVSGPAPPKASGAPGGQQRQPPQKPGPAGTTRQASQAG----- 608
QY 206 NVPTSGHPATSAHTNGPQDLRSTSGVTFRSPLVTSRARSVISVPTPRRGATQ---- 262
DB 609 FVRTGPTPTTQQRPSG-----PGPAGAP--KPQLAQKP--SQDVPPPPATAAAGGPPHPQL 660
QY 263 ----KPKPP 267
DB 661 FNLPEPAPP 669

RESULT 12
US-09-252-991A-28921
; Sequence 28921, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28921
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28921

Query Match 8.7%; Score 124.5; DB 4; Length 1034;
Best Local Similarity 24.2%; Pred. No. 0.0071;
Matches 80; Conservative 25; Mismatches 114; Indels 111; Gaps 16;
QY 42 PAGTRPT-----PDFDPFV-----EKANTR-----GVGKE 69
DB 2 PAGDPPADSRDACHRPVADPPGAGPAQAARPAASHQAQRDRRLRRQRAYPGAAVGRP 61
QY 70 ASKALGAK-----GSCETTPSRGSTPTLTPKKYKRYRISHTPSYCDSELFGRSEG--- 121
DB 62 AHRAQPARPETFDHARAPARRPRRPPAGTGRGPPPPAADAABERTRAPGHPQ 121
QY 122 ---ASFGAPRMAKGAALR---ALLWTPPTTPRGSHSP---RPREAPL----- 161
DB 122 RPPAGGAPRRARRRHAEGSCAAGWQLPQGRPGHPHPPAQARRKGAAPLRGGDRRLPA 181
QY 162 -----RAIHPAGPSKTEG-----PAADSQKLSMGLHSS-----RPLKGL 198
DB 182 PAARPRPRQPRPAGPRRHAGRSRERPAAPA-ALGRFGLLAAPAAAPAGRLCRTLAV 240
QY 199 SHSLTHLVNVPSTGH-----PATSAHTNGPQDLRSTSGVTFRS-PLVTSRAR 245
DB 241 LPAAVDAEPATGHLVRRPLAASGPARPVPATGP-----GGRGGEALRGLPL-----ER 292
QY 246 SVSISVPTSPRRGGATQKPK-----PMK 269
DB 293 PVALPGRVRRRGPGRLRRPRRPGRRSLQPWQ 322

RESULT 13
US-09-252-991A-16743
; Sequence 16743, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16743
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16743

Query Match      8.6%; Score 123.5; DB 4; Length 315;
Best Local Similarity 24.1%; Pred. No. 0.0019;
Matches 76; Conservative 29; Mismatches 100; Indels 111; Gaps 17;

```

```

QY 24 GYRVKARTSYVDETLFG--SPAGTRPTPPDFDPWVEKANTRGVGKEASKALGAKGSC 81
DB 17 GLRRPRAAATQPRAGRAPGGARRLPREL-----ATCLGPGAGRGWVVRGARP 66
QY 82 TTPSRGSTPTLTPRKKNKYPISHTPSYCDSESLFGSRSE--GASFGAPR-----MAKGA 134
DB 67 APRLPGAAPPAP---GRPAFVAATPARDEALRGDRDRELLPGTFRPRRPAGAVRGPT 123
QY 135 AKURLALLWTPP-----PTPRGSH-----GPRPREAPL--- 161
DB 124 AAGRGI---PPAGTDPAPRPRGDHRRHPQAPRHHPCLRQPRGSPRVPPQARLP 180
QY 162 ---RAIHPAGPSKTEPDAADSK-----LSMGGHLSRPLKRLGSLHSLTHLNVPSG 211
DB 181 PRRADLAAGSQAQGGGGAAGRRHAGNFCAGHGAIASRPAQ-----SG 227
QY 212 HPATSAPH-----TNG---PDLPSPSTSGVTFR-----SPLV-TSRARSVSISV 251
DB 228 APAGHLPRPRAARGRRRTGGGQQORRAVRGLIKRRAGLPAPVAGTSTA----- 280
QY 252 PSTPRGATQKXPP 267
DB 281 --TP-PGRPTAFLPP 293

```

```

RESULT 14
US-09-252-991A-17829
; Sequence 17829, Application US/09252991A
; Patent No. 8551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17829
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17829

Query Match      8.6%; Score 123.5; DB 4; Length 390;
Best Local Similarity 26.3%; Pred. No. 0.0025;
Matches 75; Conservative 17; Mismatches 108; Indels 85; Gaps 14;

```

```

QY 22 RGYRVKARTSYVDETLFGSPAGTRPTPPDFDPWVEKANTRGVGKEASKALGAKGSC 81
DB 80 RGDRLADLT-----GAQPTPL---PPGHVRLRLGGAAGAAAPATGAGARGE 124

```

```

QY 82 TTPSRGSTPTLTPRKKNKYPISHTPSYCDSESLFGSRSEG---ASFQAPRMAGDAAKLR 138
DB 125 LADSFGTAPGNRPSGRR-----TAGLRAGGIPPPASGGRRAPVRGA-FR 169
QY 139 ALLWTPPTTPRGSHSPRPRE-APLRAIHPAG-----ESKTEPGFA--ADSQKLSMG 186
DB 170 LRLSAQPVRRPRPVRPLPGGPPRARLATQPGRGSRQRPKTRPAPAGGADPAALHRG 229
QY 187 ---GLHSSRLPKRLGSLHSLTH-LNVPS-----GHP----- 213
DB 230 PGAAGRHH--RP---GTDHRRTHPRKVAATRTGTALRAAAGDFALRLCADLARRRRRSGR 284
QY 214 --ATSAPHTNGPDLRPSSTGVTFRSPLVTSRARSVSISVPSTPR 256
DB 285 RLAACTDPGGAEDLKKSQEGFSLRSISPTSPSISLADSLRPR 329

```

```

RESULT 15
US-08-317-310A-64
; Sequence 64, Application US/08317310A
; Patent No. 5858701
; GENERAL INFORMATION:
; APPLICANT: WHITE, Morris F.
; APPLICANT: SUN, Xiao Jian
; APPLICANT: PIERCE, Jacalyn H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,310A
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1321 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-310A-64

```

```

Query Match      8.6%; Score 123.5; DB 2; Length 1321;
Best Local Similarity 26.3%; Pred. No. 0.012;
Matches 72; Conservative 30; Mismatches 135; Indels 37; Gaps 11;

```

```

QY 4 PVELAVSGMOTLGLQHRCGGVYKARTSYVDETLFGSPAGTRPTPPDFDPWVEKANTR 63
DB 882 PTLISLEGLTLPSMQEYPLPTEPKSPGYINID-FGE-AGRLSPD--APPLIASAAS 937
QY 64 RGVGKEASKA--LGAKGSCETTPSRGTTPLTRKKNKYPISHTPSYCDSESLFGSR--- 118
DB 938 SLLSASSPASSIGSGTPTGTSRQSRQSLSDYMLNDFSSPKSPKPTRSGDTVSGMDGL 997
QY 119 -SEGASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGS---KTEP 174
DB 998 LSPFAS--SP-----YPLPFPSPSTSPSSLQOPL-----PPAFGLYRLPP 1036

```

Qy	175	GPAADSOQLSMGGLHSSRPLKRGLSHSLTHLNVSTGHPATGAPHTNGPQDLRPSTSGVT	234
Db	1037	ASAATSOQPTAGSSMSSEPCDNGDYTEMAFGVAATPPQPIVAPPKEGARVASP-TSGUK	1095
Qy	235	FRSPL--VTSRAKSVSISVPSTPRRGATCKXP	266
Db	1096	RLSLMDQVSGVEAFLOVSQPPDFHRCAGKVRADP	1129

Search completed: August 10, 2004, 12:14:20
Job time : 20 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 12:10:47 ; Search time 46 Seconds
(without alignments)
1834.363 Million cell updates/sec

Title: US-10-031-589-4
Perfect score: 1437
Sequence: 1 MKTPVELAVSGMQTGLQHR.....SVPTFRGGATQKPKPWK 269

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues
Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1216	84.6	229	16	US-10-452-858C-79
2	239	16.6	467	12	Sequence 1711, Ap
3	218	15.2	65	11	Sequence 8426, Ap
4	150	10.4	668	16	Sequence 134726,
5	139	9.7	429	16	Sequence 114012,
6	136.5	9.5	4019	9	Sequence 425, App
7	136.5	9.5	4019	9	Sequence 425, App
8	136.5	9.5	4019	9	Sequence 425, App
9	135	9.4	442	12	Sequence 40452, A
10	134.5	9.4	503	13	Sequence 2, Appli
11	134.5	9.4	19652	15	Sequence 7, Appli
12	134	9.3	625	16	Sequence 143626
13	134	9.3	937	16	Sequence 156445,
14	133	9.3	380	16	Sequence 200034,
15	133	9.3	514	16	Sequence 111701,

16	132.5	9.2	498	12	US-10-365-742-20	Sequence 20, Appl
17	132.5	9.2	497	16	US-10-437-963-125004	Sequence 125004,
18	132	9.2	621	16	US-10-437-963-177697	Sequence 177697,
19	131.5	9.2	277	16	US-10-437-963-177998	Sequence 177998,
20	130.5	9.1	417	16	US-10-437-963-143835	Sequence 143835,
21	130.5	9.1	878	16	US-10-722-050-2	Sequence 2, Appli
22	130	9.0	307	16	US-10-437-963-181279	Sequence 181279,
23	130	9.0	391	16	US-10-437-963-187780	Sequence 187780,
24	129.5	9.0	791	12	US-10-170-385-57	Sequence 57, Appl
25	129.5	9.0	863	15	US-10-359-012-2	Sequence 2, Appli
26	129.5	9.0	863	15	US-10-359-012-2	Sequence 14, Appl
27	129.5	9.0	19723	15	US-10-084-846A-5	Sequence 5, Appli
28	128	8.9	507	13	US-10-078-547-24	Sequence 24, Appl
29	128	8.9	514	16	US-10-437-963-195781	Sequence 195781,
30	128	8.9	797	14	US-10-156-761-10907	Sequence 10907, A
31	127.5	8.9	388	16	US-10-437-963-160033	Sequence 160033,
32	127.5	8.9	713	16	US-10-437-963-150342	Sequence 150342,
33	127.5	8.9	1870	16	US-10-408-765A-120	Sequence 120, App
34	127	8.8	238	16	US-10-437-963-145797	Sequence 145797,
35	127	8.8	481	16	US-10-437-963-176449	Sequence 176449,
36	126.5	8.8	1321	16	US-10-694-874-4	Sequence 4, Appli
37	126	8.8	695	14	US-10-122-805-4	Sequence 4, Appli
38	125.5	8.7	699	14	US-10-121-988-143	Sequence 143, App
39	125.5	8.7	699	14	US-10-300-562-143	Sequence 143, App
40	125.5	8.7	699	14	US-10-237-551-143	Sequence 143, App
41	125.5	8.7	699	14	US-10-237-551-254	Sequence 254, App
42	125.5	8.7	1103	16	US-10-437-963-166325	Sequence 166325,
43	125	8.7	469	16	US-10-437-963-136216	Sequence 136216,
44	124.5	8.7	365	16	US-10-437-963-116009	Sequence 116009,
45	124.5	8.7	556	16	US-10-437-963-137918	Sequence 137918,

ALIGNMENTS

RESULT 1
US-10-452-858C-79
; Sequence 79, Application US/10452858C
; Publication No. US20040086945A1
; GENERAL INFORMATION:
; APPLICANT: Sreekrishna, Kotikanyadanam
; APPLICANT: Gerwe, Gina S
; APPLICANT: Toerner, Daniel R.
; TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THERE
; FILE REFERENCE: 8956P
; CURRENT APPLICATION NUMBER: US/10/452,858C
; CURRENT FILING DATE: 2003-06-02
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (217)..(217)
; OTHER INFORMATION: The 'Xaa' at location 217 stands for Lys, Arg, Thr, or Met.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (219)..(219)
; OTHER INFORMATION: The 'Xaa' at location 219 stands for Tyr, Cys, Ser, or Phe.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (225)..(225)
; OTHER INFORMATION: The 'Xaa' at location 225 stands for Leu, or Phe.
US-10-452-858C-79

Query Match 84.6%; Score 1216; DB 16; Length 229;
Best Local Similarity 98.7%; Pred. No. 2e-84;
Matches 226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 11 GMDTGLQHRGGRVYKARTSYDETUGSAGTRPTFPDPDPWKEKANTRGVKEA 70
|||||

```
Db 1 GMOTLGLQHRGGRVYKARTSYVDLTFGSPAGTRTPDPDPFWVEKANRTGKGEA 60
QY 71 SKALGAKGSCETTPSRGSTPTLTTRKKNKYRPISTHTPSYCDLSFGSRSEGSFGAPRMA 130
Db 61 SKALGAKGSCETTPSRGSTPTLTTRKKNKYRPISTHTPSYCDLSFGSRSEGSFGAPRMA 120
QY 131 KGDAAKRALWTPPTPRGSHSPRPREALRAIHAPGSKTEFGPAADSKQLSMGLHS 190
Db 121 KGDAAKRALWTPPTPRGSHSPRPREALRAIHAPGSKTEFGPAADSKQLSMGLHS 180
QY 191 SRPLKRGLSHLTHLVNSTCHPATSAHNTNGPDLRPSGVTFRSPL 239
Db 181 SRPLKRGLSHLTHLVNSTCHPATSAHNTNGPDLRPSGVTFRSPL 229

RESULT 2
US-10-276-774-1711
; Sequence 1711, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Tang, Y, Tom et al
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1711
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1)...(467)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-1711

Query Match 16.6%; Score 239; DB 12; Length 467;
Best Local Similarity 29.4%; Pred. No. 5.6e-10;
Matches 94; Conservative 30; Mismatches 94; Indels 102; Gaps 17;

QY 5 VELAVSGMOTLGLQHR--CRGGYR-----VKARTSYVDLTFGSPAGTRTPDPDPFW 56
Db 176 MRLCLAAQAAGLPHRTSIRPGWRRLTEPEAWARR---HRRPWGQGVAVRPPQGAAPP 232
QY 57 VEKANRTGKGEASKALGAKGSCETTPSRGSPPTLTTRKKNKYRPISTHTPSYC---DBS 113
Db 233 SHQGRRT-----NTDPS--ATPRLT-----VNGRCLAPDLK 261
QY 114 LFGSRSEGSFGAPRMAKGAAXLALLWTPPTPRGSHSPRPREALRAIHAPGSKTE 173
Db 262 APASGPRGWRMRGMPQS-----GALLWTPPTPRGSHSPRPREALRAIHAPGSKSR 314
QY 174 PGPAADSKQLSMGLHSRPLKKG-----LHSL-----THLVNSTCHPAT 216
Db 315 AGASGRLEVIYGNWTLFTPEAGTFLIPSTXMSPALVIQPPVPTQGLSGLRQ 374
QY 217 APTNGPDLRPSGVTFR-----RSLPVTSRARSVISVPSTPR----- 257
Db 375 XP--SGAFWXL--PGLAQIAFQCHLPHDVEGVPNRNQSLGND---TLSSGLPWGRQVWPL 429
QY 258 ---GG--ATQKP-----KPPW 268
Db 430 ARVGGHSPREPQVLKKPLW 449

RESULT 3
US-09-864-408A-8426
```

```
; Sequence 8426, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; FILE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Enco
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8426
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-8426

Query Match 15.2%; Score 218; DB 11; Length 65;
Best Local Similarity 71.4%; Pred. No. 2.4e-09;
Matches 40; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 12 MDTLGLQHRGGRVYKARTSYVDLTFGSPAGTRTPDPDPFWVEKANRTGKVG 67
Db 1 MKAHLQHRSTSYRVKARASYVDLTFGSPARTPAQDPDFPWWQNCNRSRGVG 56

RESULT 4
US-10-437-963-134726
; Sequence 134726, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 134726
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36471C.1.pep
US-10-437-963-134726

Query Match 10.4%; Score 150; DB 16; Length 668;
Best Local Similarity 26.7%; Pred. No. 0.0047;
Matches 72; Conservative 21; Mismatches 113; Indels 64; Gaps 13;

QY 23 GGVYKARTSYVDLTFGSPAGTRTPDPDPFWVEKANRTGKGEASKALGAKGSC 81
Db 80 GGVNPPSPSITGTPPTGGGGYVTPPTSDTPPSPSDTSPSTPGGCSSTP-----PCD 134
QY 82 TTSPRGS--TTLTPRKNKYRPISTHTPSYCDLSFGSRSEGSFGAPRMAKGAAXLAL 140
Db 135 APPSPSDTSPSTPGGGYSP---TFS-----DTPPSPSD----- 168
QY 141 LWTPTPTP---RGSHPSPRPREAP-----LRAIHAPG-----PSKTEPGPAADSKQLSM 185
Db 169 --TSPTTGGGGYVTPPTSDAPPSPSDTSPSTPGGGYVTPPTSDAPPSPSDTSPPT 226
QY 186 GGLHSSRPLKGLSHSLTHLVNSTCHPATSAHNTNGPDLRPS-----GVTR 236
```

```

Db 227 GCGGGYTPSDTPSPSSGSPPTT--PGGGGYTPTSDTPSPSSGSSRTTPGCGSTP 284
QY 237 SLVTSRARSVSISVPSPTRRGATQKPKP 266
Db 285 TFCGTPAPSSGTS-PTTP--CGSYYPPTP 311

RESULT 5
US-10-437-963-114012
; Sequence 114012, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114012
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(429)
; OTHER INFORMATION: unsure at all xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17745C.1.pep
; OTHER INFORMATION:
US-10-437-963-114012

Query Match 9.7%; Score 139; DB 16; Length 429;
Best Local Similarity 24.5%; Pred. No. 0.02;
Matches 83; Conservative 36; Mismatches 100; Indels 120; Gaps 19;

QY 21 CRGGYR---VKARTSVYDETIFGS-----PAG-----TRPTPD-----FDPWPVEKAN 61
Db 5 CTGVRPWFVRVRRRYTHRQLLTQSRPHPPPAAPLVPSQPPPPDAPLPIRPPSIGCAS 64
QY 62 RTRGVGXEASKALCAGKS-----CETTPSRG-----STPTLTTRKKNKY 100
Db 65 RAKPASISAAASPAEFAAIGCTSPSRPIPPENDAMPARCNELPRPLAVRGRHAS 124
QY 101 RPISH-----TPSYCDESIFGSRSEGASFGAPRAKDAAKLRAALLWTPPTPRGSH 152
Db 125 NDRSHHRRSRRCCQMPSHLAAILTS-SAAATXRSRPS-QAVRCRSL---QPAP--- 175
QY 153 SPRPREAPLRAIHAPGSKTEPGPAADSQKLSMGGLH-----SSRPLKRGLSH- 200
Db 176 -----AVAAELHVAGDSILPPP-----HVAGLHDACAGLAHAARPPRCIDHH 223
QY 201 -----SULHNVPTGHPATSA---PHNNGQDLRPSGTGVTFRSPL---VTSRAR----- 245
Db 224 GRCLASQSPQSMPPQPPVTAAGVSHNRQPDL---TRGVVASSCIAGVTGSRXPPPP 280
QY 246 -----SVSISVPSPTRRRGATQKPKP 267
Db 281 HLSSGRASVLGTGSGSGSVSRXVATAPGR-----RRPAPP 315

RESULT 6
US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
```

```

; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather.
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-738-973-425

Query Match 9.5%; Score 136.5; DB 9; Length 4019;
Best Local Similarity 24.0%; Pred. No. 0.38;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;

QY 29 ARTSVYDETIFGSPAGTR-----PTPDPFDPWPVEK-----ANTRGVGK 68
Db 988 SRPLQMNETTANRPSVRDLGSSSTTNDPYAKPDPTRPVMTDQFPKSLGSLRSPVSE 1047
QY 69 EASKALGAKGSC-----ETTPSRGSTTLTPRKNK-----YRPISTHTPS 108
Db 1048 QTAKGPIAAGTSDHFTKPSRADVFQRIQIPDSYARLLITPAIDSGFGPKTQMPPPS 1107
QY 109 YCDESIFGSRSEGASFGAPRAKDAAKLRAALLWTPPTPRGSH--SPRPREAPLRAIHP 166
Db 1108 SQDP--YGVVSQ-----ASRRLSVDPYERPAL--TPRIDNFHNSQNDPYSQPLTGP 1158
QY 167 -----APSK--TEPG-----PAADSQKLSMGGLHS-----SRPL 194
Db 1159 AVNESFAHPSRAFSQPGTISRPTSDQPSQPPGTPRPFVVDYSQSSGTPARNTDPYSQP- 1217
QY 195 KEGLSHLTHLNVPTGHPATSAHTNGQDLRPSGTGVTFRSPLVTSRARSVSISVPS 254
Db 1218 -----POTPRPTTVDPYSQQPQTRPRSTQTDLFVTPYTNQHSDDPYAHPPT 1264
QY 255 PRGGATQKPKP 267
Db 1265 PRGGSVPSYQPP 1277

RESULT 7
US-09-854-133-425
; Sequence 425, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
```


; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/968,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Human
US-10-078-547-2

Query Match
Best Local Similarity 25.8%; Pred. No. 0.052;
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY 42 PAGTR-PTPPDFPPVWEKANRTRGVGKEASKALGAKGSCETTPSRGSTPT-LTPRKN 98
Db 207 PGGRQPSGPTPPPGNREGTALGGSIQSPSSSPFNRPPLPTPSRALDDKPPP 266

QY 99 KYRPISTPSYCDSESLFGSRSESGASFCAPRMAGDAKLRALLWTPPTPRGSHSPRPR- 157
Db 267 PPPPVGNRPDSIHREAV-----PPPPQNNKPPVPST 297

QY 158 ---EAPLRA-IHPAGSKTEPGPAADSKLSMGLHSSRLPKRGLSHSLHLNVPSTGH- 212
Db 298 PRFSAPHRPLRPPPSRPGPLPPS-----SSNDETPLRQNLSSSTPLPSPGRS 354

QY 213 -PATSAPTNGPDLR--PSTSGVTRSPLVTSRAASVSISVSTPR---RGATOK--- 263
Db 355 GPLPPPSRPPPPVVRDPPGRSG-PLPPPPVSRNGSTSRALPATQLPSRSGVDSPRSG 413

QY 264 PKSP 267
Db 414 PRPP 417

RESULT 11
US-10-084-846A-7
; Sequence 7, Application US/10084846A
; Publication No. US2004006028A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 7
; LENGTH: 19652
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-7

Query Match
Best Local Similarity 25.7%; Pred. No. 3.3;
Matches 80; Conservative 29; Mismatches 103; Indels 99; Gaps 16;

QY 26 RVKARTSYVDETLFGSPAG-----TRTPDFDPVWEKANRTRGVGKEASKAL 74
Db 11238 RRPVQPAHVDAEAGLGLGVVFDGVDVPSRSVIPSELPOWVER---EQSETEPMAS 11293
QY 75 CAKGSC-----ETTPSRGSTPTLTP-----RKNKYRP--- 102
Db 11294 SSQGRCTWRSPAQVASKWSPANSIDSTSASRST-LVTPCTMTNSARRRFSRPSRA 11352
QY 103 -ISHTPSYCDSESLFGSRSESGASFCAPRMAGDAK-LRALLW-----TPP 145
Db 11353 RIS-TPG-CPDIARVSAPSSAPSSRSTRAPGNAKNAKASRSGWGLSRSEAVAYRTPL 11410
QY 146 PTPRGSH--SPRPREAPLRAIHPAGPSKTEPGPAADSKLSMGLHSSRLPKRGLSHSLT 203
Db 11411 YTKSFTGSSSRTRRRISRSFTVALP-----PLSGALPSETAWAGDHRSAPLSRGATH--- 11463
QY 204 HLNVPSTGHPATSAP-----HTNGPOD-----LRPSTSGVTRSPLVTSRAVS 247
Db 11464 -----CAHSTRPFPPIISKEERSAGPTEAHAARVSPSASGVSAQSP-KTSRASSV 11516
QY 248 SISVSTPRRG 258
Db 11517 AVATITSSRAG 11527

RESULT 12
US-10-437-963-143626
; Sequence 143626, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 143626
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (625)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_44516C.1.pep
US-10-437-963-143626

Query Match
Best Local Similarity 23.5%; Pred. No. 0.072;
Matches 73; Conservative 27; Mismatches 103; Indels 108; Gaps 13;

QY 41 SPAGTRPTP-----PDFDPVWEKANRTRGVGKEASKALGAKGSCETTPSRGSTPTL 92
Db 227 APATASTXPPHRPPHRLPXCSP-REAAARPPPLSLDLAVALLPSGSAARPPR----- 279
QY 93 TPRKNKYRPISTPSYCDSESLFGSRSESGASFCAPRMAGDAKLR----- 138
Db 280 -----REVAFTPPWSDSASARPO-----PREGPAAARLAATARGPQASRSPG 325
QY 139 -----ALLWTPPTPRGSHS-----PRPREAPLRAIHPAGPSK 171
Db 326 PRRVVRPPQAQDKGAILITTPPTPMQASAGBGRVVRVPPRPRTRGSPSSPHLPPCR 385

QY 172 -----TERGPAADSKLNGGLHSSRLPKRGLSHSLTHLNV-----STGHPATSA 217
 Db 386 FGRPACPTLLLSPPSPASISLAASVA---CPAPSPSTLPRTRDPSPPXSTAGAPPSTA 442
 QY 218 --PHNGQDLRPSTSGVTFPSPLVTSRAR-----SVSISV-----PST-PR 256
 Db 443 SPPPLISAASLRRSPSPSPPLVTTISRRQSTSPSPPPPSISLAASVACPAPSPSTLPR 502
 QY 257 RGGATQKPKPP 267
 Db 503 RTRPSSPPSP 513

RESULT 13

US-10-437-963-156445
 ; Sequence 156445, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 156445
 ; LENGTH: 937
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(937)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_56111C.1.pep
 US-10-437-963-156445

Query Match 9.3%; Score 134; DB 16; Length 937;
 Best Local Similarity 26.2%; Pred. No. 0.11;
 Matches 68; Conservative 21; Mismatches 83; Indels 88; Gaps 14;
 QY 40 GSPAG-TRPTPPDPDPFWVEKANRTGKVGKEASKALGAKGSCETTPSRG-----88
 Db 702 GPPPGSRPTQP---PEPLSRVTR-----VDAARPTASRASPPHLLSPS 747
 QY 89 -----TPTLTRKKKRYPISTPSYCDSELF-GSRGEGASFGAPRMAKGAAXL 137
 Db 748 SLSRAAGRLQPLSPR-PSAARPCRLFRRRRLFPAGAAANSWNHRRRLVAGTKVAL 806
 QY 138 RALLWTPPTTPRGSHGFRPRE-APLR---AIHPAGPS-----KTEPGAADSKLSNG--G 187
 Db 807 PPLYCQPLPFRSGAGRLRLPFRRRVALHGLSVLPERRP-----RRLSLGLVG 860
 QY 188 LHSRLKRLGLSHLTHLNVPTSGHATSPHTNGPDLPSTSGVTFPSPLVTSRARSV 247
 Db 861 THSGSPFR-----ASAPADVAVSTGIA-----AGKV 888
 QY 248 SISV---PSTPRRGATQKP 264
 Db 889 VLGIASPPSKFLRGPSSPP 908

RESULT 14

US-10-437-963-200034
 ; Sequence 200034, Application US/10437963
 ; Publication No. US20040123343A1

; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 200034
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(380)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_95543C.1.pep
 US-10-437-963-200034

Query Match 9.3%; Score 133; DB 16; Length 380;
 Best Local Similarity 24.2%; Pred. No. 0.049;
 Matches 62; Conservative 16; Mismatches 114; Indels 64; Gaps 9;
 QY 41 SPAGTRPTPPDPFWVEKANRTG-----VGKEASKALGAKGSCETTPSRG 87
 Db 86 SAAAAAPPTAAPPHRRRCRCGNPPPLPSPSSAASAPSRARSPPAATPCPARP 145
 QY 88 STPTLTRKKKRYPISTPSYCDSELFSGRSEGAFCAPR---MAKGAAXLALLWT 143
 Db 146 RCGSATP--PAPWRXSPSPSPAPAGCAARTPPPAYVHRLSMITPPPRHHAAL 203
 QY 144 PPPTPR---GSHSPRPREAPLRAIH-----PAGPSKTEPGAADSKLSMGLHSSRP 193
 Db 204 PPAPPRPNRTTRPARRRHRLIRIPPPPLPAPPSPAPPGP-----246
 QY 194 LKRLGLSHLTHLNVPTSGHATSPHTNGP--ODLRPSTSGVTFPSPLVTSRARSVSISV 251
 Db 247 -----RHPPFPSP-HPPXPCGPRRPPAGRNHLHSPPPPNPPLPFSR-----GP 290
 QY 252 PSTPRRGATQKPKPP 267
 Db 291 PPSPSRSPATDLPPAP 306

RESULT 15

US-10-437-963-111701
 ; Sequence 111701, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 111701
 ; LENGTH: 514

```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(514)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_15655C.1.pap
; US-10-437-963-111701

Query Match          9.3%; Score 133; DB 16; Length 514;
Best Local Similarity 26.6%; Pred. No. 0.069;
Matches 64; Conservative 20; Mismatches 117; Indels 40; Gaps 9;

QY  .42  PACTRTPDPPDPWVEKANRTGKVEAS-----KALGAKGSCETTPSRGSTPTLTP 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239  PARTLPAPPXADPPGRGXPRRRATGTRAAPPFACFTRRSFAGKGCWSPRPPPGGSFTAP 298

QY  95  RKCKKYRPISHTFSYCDLSFGSRSEGSFGAPRMAKGDAAKLRALLWTPPTPRGSHSP 154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299  -----PPPPPPRGAPRAAASAAASAAAGRAPRAPRPPRPPPPPSARASSSP 352

QY  155  RPREAFLRA-----IHPAGPSKTEPGPAADSQKLSMGGLHSSRPLKKGLSHLTHLNVST 210
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353  RHSATPAPAHXPRPHSPGCGRPPTTRAPSS-----FWPPPTGAAP-----RGPGT 398

QY  211  GHFATGAPHTNG-PQDLR---PSTSGVTFRSPLVTSRARSVSISVSTPRRGATOKPKP 266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399  FAPSSSWPPPPAGTFAALRWAAAPASAPGTACS--AAGRARG---SSPPPPQHRHHHPSPPP 453

QY  267  P 267
Db 454  P 454
```

Search completed: August 10, 2004, 12:13:50
Job time : 47 secs

This Page Blank (uspto)

LTR gag/pol polypr
toucan gene protei
calcium channel pr
hypothetical prote
synapsin Ia - rat
EBNA-LP protein -
AP4 protein - mous
nascent polypeptid
tpr homolog - fui
hypothetical prote
hypothetical prote
DNA-binding protei
extensin-like cell
proline-rich cell
hydroxyproline-ric
BPLF1 protein - hu

30 115.5 8.0 1456 2 T01397
31 115.5 8.0 2176 2 T13806
32 115.5 8.0 2212 2 A41098
33 115 8.0 380 2 D70516
34 115 8.0 704 2 A30411
35 114.5 8.0 308 2 S77938
36 114.5 8.0 1211 2 T42230
37 114.5 8.0 2187 2 T30828
38 114 7.9 2346 2 T13829
39 113.5 7.9 346 2 T46916
40 113 7.9 1952 2 T48814
41 112.5 7.8 2282 2 T42717
42 112 7.8 214 2 T10737
43 112 7.8 214 2 T09854
44 112 7.8 620 2 S06733
45 112 7.8 3149 1 Q08E8

ALIGNMENTS

RESULT 1

T28770
hypothetical protein W03D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T28770
R:Kohling, F.; Wohlmann, P.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of C. elegans cosmid W03D2.
A:Reference number: Z20519
A:Accession: T28770
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-539 <ROH>
A:Cross-references: EMBL:AF000299; PIDN:AA048255.1; GSPDB:GN00022; CBSP:W03D2.1
A:Experimental source: strain Bristol N2; clone W03D2
C:Genetics:
A:Gene: CBSP.W03D2.1
A:Map position: 4
A:Introns: 40/3; 88/3; 115/3; 146/3; 173/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Result No.	Score	Query Match	Length	DB ID	Description
1	144.5	10.1	539	2 T28770	hypothetical prote
2	136	9.5	2715	2 T13049	eyelid - fruit fly
3	132.5	9.2	580	2 T43481	probable mucin DKF
4	130.5	9.1	705	2 A35363	synapsin I splice
5	128	8.9	403	2 S52796	ppl2 protein - hu
6	127.5	8.9	1870	2 S37671	MHC class III hist
7	127.5	8.9	1872	2 S36152	MHC class III hist
8	127	8.8	302	2 T32711	hypothetical prote
9	126.5	8.8	1199	2 A40670	nuclear envelope p
10	126	8.8	706	2 B30411	synapsin Ia - bovi
11	125.5	8.7	699	2 C43674	US4 protein - huma
12	124	8.6	576	2 T36729	probable serine/th
13	123.5	8.6	724	2 T47149	hypothetical prote
14	123	8.6	886	2 S29605	glycoprotein 350/2
15	122.5	8.5	907	1 Q08E21	membrane antigen 9
16	122	8.5	1777	2 T34369	hypothetical prote
17	121	8.4	721	2 E70766	high molecular mas
18	121	8.4	1151	2 T18535	MHC class III hist
19	120.5	8.4	2142	2 B35098	NIMA-like protein
20	120	8.4	779	2 A37177	unconventional myo
21	120	8.4	3530	2 A92266	immediate-early pr
22	119	8.3	1446	1 A45344	serine-rich protei
23	118.5	8.2	1077	2 A44067	En/spm-like transp
24	116.5	8.1	306	2 T52305	unconventional myo
25	116.5	8.1	3511	2 A59295	hypothetical prote
26	116	8.1	635	2 F15477	hypothetical prote
27	115.5	8.0	846	2 T21700	hypothetical prote
28	115.5	8.0	883	2 A96662	hypothetical prote
29	115.5	5.0	1285	2 T14171	ataxin-2 - mouse

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 10, 2004, 12:08:42 ; Search time 17 seconds
(without alignments)
1522.089 Million cell updates/sec

Title: US-10-031-589-4
Perfect score: 1437
Sequence: 1 MKTPVELAVSGMOTLGLQHR.....SVFSTPRRGATQKPKPPWK 269

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

A:Residues: 262-580 <POU2>
A:Cross-references: ENBL:AL117491; NID:g5911958; PIDN:CAB55954.1; PID:g5911959
A:Experimental source: adult testis; clone DKFZp434B061
C:Genetics:
A>Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

Query Match 9.2%; Score 132.5; DB 2; Length 580;
Best Local Similarity 26.3%; Pred. No. 0.14;
Matches 77; Conservative 30; Mismatches 125; Indels 61; Gaps 17;

QY 3 TPVELAVSGMOTLGLQHRCRGGVKARTSV-----DITLFSGPAGTRPT--PPDFDPW 56
DB :|::||:|||||
QY 57 VEKANRTRVGVEKASKALKAGKCETTPSRGSTTLTPRKKNKYRPISHTPSYCODESLFG 116
DB :|::||:|||||
QY 99 PSRASTPRRLFRASPMGSPHRASPMRTTTPRAS-PGTTFSTASP---TGTPS--SASPTG 151
DB :|::||:|||||
QY 117 SRSGACGFCAAPMA----KGDAAKL-----RALLTWPPTPRGSHSPREAPLAHPA 167
DB :|::||:|||||
QY 152 TTPRASFTGGTPRAWATRSSTASLTTPRASLTRMPPRASPTRIP-PRESP-RMSHRA 209
DB :|::||:|||||
QY 168 GPSKETP-GPAADSQKLMSGHLSSPLKEGL-SHSLTHLVNPSTGHPATSAFHTNGP 223
DB :|::||:|||||
QY 210 SPRTTTPRASPTRRRPPRAS----PTRTPPRESLTSTRASPTRMPPRASPTRERP----- 260
DB :|::||:|||||
QY 224 QDLRPSNSGYTER-SPLVTSRASVSLSVSTPRGGATCKPK-----PPW 268
DB :|::||:|||||
QY . 261 ---RASDTGGPPRASMTTPPRAS-----PTTPRASFTTTTPRASLTRTESW 304
DB :|::||:|||||

RESULT 4
A35363
synapsin I splice form a - human
C:Species: Homo sapiens (man)
C>Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 01-Dec-2000
C:Accession: A35363; B35363; A35805
R:Suedhof, T.C.
J. Biol. Chem. 265, 7849-7852, 1990
A:title: The structure of the human synapsin I gene and protein.
A:Reference number: A35363; PMID:90243651; PMID:2110562
A:Accession: A35363
A:Molecule type: DNA
A:Residues: 1-705 <SUE>
A:Cross-references: GB:M58371; GB:J05431
A:Accession: B35363
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-659, 'XASPAQAQP' <SU2>
A:Cross-references: GB:M58378; GB:J05431
R:Sauerwald, A.; Hoeseke, C.; Oschwald, R.; Kilimann, M.W.
J. Biol. Chem. 265, 14932-14937, 1990
A:title: The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-less
A:Reference number: A35805; PMID:90368667; PMID:2118519
A:Accession: A35805
A:Molecule type: DNA
A:Residues: 1-125 <SAU>
A:Cross-references: GB:M55301; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630
C:Genetics:
A:Gene: GDB:SYN1
A:Cross-references: GDB:l19606; OMIM:313440
A:Map position: Xp11.23-Xp11.23
C:Keywords: actin binding; alternative splicing; phosphoprotein

Query Match 9.1%; Score 130.5; DB 2; Length 705;
Best Local Similarity 23.5%; Pred. No. 0.23;
Matches 61; Conservative 24; Mismatches 108; Indels 67; Gaps 9;

QY 40 GSPACTRTPPDPDFPWTEKANTRTG-----VGKEASKALKAGKGCCTTPSRGSTP 90
DB :|::||:|||||
QY 454 GPFAQQR-PPQGPPPGGPGQRGPQQPQQHQQLSLGSLGPAG-SLPFLQSLPSP 511
DB :|::||:|||||
QY 91 TLTPKKNKRYDISHTPSYCDSLFSGRSEGAFCAPRNAMKDAAKLRALLWTPTPTPR- 149

	QY	80	CETTPSR---	GSTPTLTTRKKNKVRPI--SHTPSYCDESLFGSRSGASFGAFPRMAKGDA	133
	Dd	763	PVDPKLAVNGDVFTATPAEP---RPLTSPLEQAEDDDCKGRSE-	----	824
	QY	136	KLRALLWTPTPTPGSHSPREAPLRAIHGAPGSKTTEPGAADSQKLSMG-	-----	187
	Dd	825	-----TPVPV-----PPP-----PYLASYPGFENGTPGPISRFPLEEPPRLPWPP	968	
	QY	188	-----LHRSRPLKRGLSHLTHLVNSTGHFATSAPHGTNGPQDLRPSTSGVTFRSPLV	240	
	Dd	869	GSDEVAKIQTTPPKKEPPEKEETAQLTGPEAGRKLPASRGAGPPPPRESRETETRWGPRP	928	
	QY	241	TSRARSVSISVP-STPRGGATQKPKPPWK	269	
	Dd	929	GSSRRGIIPBEEFGAPPERRAGPIKKDPPPTX	958	

RESULT 8

T32711
 hypothetical protein T22D1.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T32711
 R:Geisel, C.; Bradshaw, H.; Hawkins, M.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid T22D1.
 A:Reference number: Z21211
 A:Accession: T32711
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-302 <GEI>
 A:Cross-references: EMBL:AF039052; PIDN:AA94271.1; GSPDB:GN00022; CESP:T22D1.2
 A:Experimental source: strain Bristol N2; clone T22D1
 C:Genetics:
 A:Gene: CESP:T22D1.2
 A:Map position: 4
 A:Introns: 99/1
 C:Superfamily: proline-rich protein

Query Match 8.8%; Score 127; DB 2; Length 302;
 Best Local Similarity 24.6%; Pred. No. 0.17;
 Matches 65; Conservative 15; Mismatches 98; Indels 86; Gaps 11;
 QY 44 GTRPTPPDFDPP--WVEKANTR-----GVGKEASKALGAKG--SCETTPSRGSGTTL 92
 Db GTTPPPPTGEPQDLSAEGNASRRPPPPKGTGTPPPPTGEPQDLSGEGNASRRPP-- 129
 QY 93 TPRKKNKVRPISTHPSYCDLSFGRSEGASFGAPRVAKGDAKRLALLTPPTPRGSH 152
 Db 130 -PPKGTGSPPPPT-----GEPQDLSGEGNASR-----RPPPPPKGTG 166
 QY 153 SPRP-----REAPLRAHPAGSKTEPGPAADSOXLSMGLHSSRPLK 195
 Db 167 SPPPTGEPQDLSGEGNASRRPP-----PPKGTGTPPPPTGEPQDLSAEGVSRP-- 220
 QY 196 RGLSHLTHLNVPTGHPATSAHTNGPDLR-----PSTGVTFRSLVTSRAR 245
 Db 221 -----PPPKGTGSGTPPTGEPQDLSGEGNASRRPPPPKGTGSGPPPTGEPQ 270
 QY 246 SVSISVPTPRGGATOKPKPPWK 269
 Db 271 DLS-----GEGNASRRPPPPK 287

RESULT 9

A40670
 nuclear envelope protein POM 121 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 05-Nov-1999
 C:Accession: A40670
 R:Hallberg, E.; Wozniak, R.W.; Blobel, G.
 J. Cell Biol. 122, 513-521, 1993
 A:Title: An integral membrane protein of the pore membrane domain of the nuclear envelope
 A:Reference number: A40670; MUID:93328754; PMID:8335683
 A:Accession: A40670
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1199 <HAL>
 A:Cross-references: GB:Z21513; NID:G396746; PIDN:CAA79725.1; PID:G396747
 F:803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X-F-

Query Match 8.8%; Score 126.5; DB 2; Length 1199;
 Best Local Similarity 26.1%; Pred. No. 0.74;
 Matches 67; Conservative 26; Mismatches 131; Indels 33; Gaps 11;
 QY 28 KARTSYVDETLF---GSPAGTRFTPPDFPPWVEKANTRGVGKEASKALGAKGSCETT 83
 Db 362 RSTSSVSLTSTCTGIPSSRNA-----ITSSYSSTRGV-SQLWKRSGTSPSPSS 413
 QY 84 P--SRGSTPLTPRKKNKVRPISTHPSYCDLSFGRSEGASFGAPRVAKGDAKRLALL 141

Db 414 PASSRSQTPE-RPAKTRREEEFCHOSSSSAPLVTKESPGKEKVTDPATGKQOS-----L 466
 QY 142 WTPPTTPRGSHSPRPREAPL-----RAIHPAGSKTEPGPAADSOXLSMGLHSSRPLKRG 197
 Db 467 WTSPTTP-GSSGORKKIKQLPSRRGDQLTLPPPELGVISITAEOLDMERRASLQWENKV 525
 QY 198 LSHSLTHLNVPTG-HPATSAHTNGPDLRSTSGVTF-----RSPLVTSRARSVISI 252
 Db 526 LEDXTDDASTPATDTSPTSPFTTLTPTVGASPASLPAPSSNPPLLESLLKMGQSPAP 585
 QY 253 ST--PRRGATOKPKPP 267
 Db 586 SSSPEPEAAATVAAPSP 602
 RESULT 10
 E30411
 synapsin Ia - bovine
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 01-Dec-2000
 C:Accession: E30411; F30411; A35758
 R:Suedhof, T.C.; Czernik, A.J.; Kao, H.T.; Takei, K.; Johnston, P.A.; Horiuchi, A.; Kana
 Science 245, 1474-1480, 1989
 A:Title: Synapsins: mosaics of shared and individual domains in a family of synaptic ves
 A:Reference number: A30411; MUID:89388265; PMID:2506642
 A:Accession: E30411
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-706 <SUB>
 A:Accession: F30411
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-661, 'KA', 664, 'PAQAQ', 'SU2'
 R:Hall, F.L.; Mitchell, J.P.; Vulliamt, P.R.
 J. Biol. Chem. 265, 6944-6948, 1990
 A:Title: Phosphorylation of synapsin I at a novel site by proline-directed protein kinas
 A:Reference number: A35758; MUID:90216728; PMID:2108963
 A:Accession: A35758
 A:Molecule type: protein
 A:Residues: 532-556 <HAL>
 C:Keywords: actin binding; alternative splicing; phosphoprotein
 F:551/Binding site: phosphate (Ser) (covalent) (by proline-directed kinase) #status expe
 Query Match 8.8%; Score 126; DB 2; Length 706;
 Best Local Similarity 22.9%; Pred. No. 0.46;
 Matches 60; Conservative 23; Mismatches 87; Indels 92; Gaps 12;
 QY 40 GSPAGTRPTPPDFPPWVEKANTRG-----VGEASKALGAKGSCETTPSRGSP 90
 Db 454 GPPAQARP-PPQGGPFGPGPQPGPPLQORPTTQCGHLSGLGPPAG-SPLFQRLPSP 511
 QY 91 TLTPRKKNKVRPISTHPSYCDLSFGRSEGASFGAPRVAKGDAKRLALLWTP--PPTP 148
 Db 512 TSVFQ-----QPASQ-----ATPTQGGQGRSRPVAGGFGAPPAT 546
 QY 149 RGSHPSPR-----EAPLRA--IHPA-----GPSKTEPGPAADSOXLSM 185
 Db 547 RPPASPSFQAGPPQATRTQTSVSGOAPPKASGVPPGGQORQGPQKPPGPGAGTRQASQ 606
 QY 186 GGLHSSRPLKRLSHLSLTHLNVSTGHPATSAHTNGPDLRSTSGVTSRSLVTSRAR 245
 Db 607 AG-----PMRTGPTTQQRPPSGPGAGPTKQLAKP----- 641
 QY 246 SVSISVPTPRRGATOKPKPP 267
 Db 642 SQDVPPATAAAG-----PP 657
 RESULT 11
 C43674
 US4 protein - human herpesvirus 2 (strain HG52)
 C:Species: human herpesvirus 2

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-2000

C;Accession: C43674
R;McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.
J. Gen. Virol. 68, 19-38, 1987
A;Title: DNA sequence and genetic content of the HindIII 1 region in the short unique co-
utary comparisons.
A;Reference number: A43674; MUID:87111457; PMID:3027242
A;Accession: C43674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-699 <MCG>
A;Cross-references: EMBL:X04798; NID:g59900; PIDN:CAA28490.1; PID:g59904
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 8.7%; Score 125.5; DB 2; Length 699;

Best Local Similarity 23.0%; Pred. No. 0.49; 92; Indels 79; Gaps 11;

Matches 59; Conservative 27; Mismatches 92; Indels 79; Gaps 11;

QY 35 DETLFGSPAGTRTPDPDPFPPWVEKANTRGVGKEASKALGAKGSCETTPSRGSTPTLTP 94

DB 405 EETAVASPPATASVE---SSPLPAAAAATPGAGHTNTSSASAAKTPTTPA---PTTPP 457

QY 95 RKKNKYRPSHTSYCODESLFGSRSEGFAGPRMAKGDAAKRALLLWTPP---PTPRG 150

DB 458 PTSTHATPRTPP-----GPQT-----TPCPATPGPYG 486

QY 151 SHSPRPEALRAIHAPGPKSTPPGPAADSOKLSMGLHSSRLKGLSHS-----LTHL 205

DB 487 ASAPRADSLTASPPA---TAPGSAANVSA---ATTATPGTGARTPTDPKTHP 539

QY 206 NVPTSTGHATSAP---HTNGPDLR-----PSTSGVTFRSPVLTSRARSVSIS 250

DB 540 HGPADAPPGFAPPPPEHRGPEEFEGAGDEPDDSATGLAFRTPNPK-----591

QY 251 VPSTPRGGATQPKPP 267

DB 592 --PPPARPGPIRPTLPP 606

RESULT 12

T36729

Probable serine/threonine-specific protein kinase (EC 2.7.1.1) - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Jul-2001

C;Accession: T36729

R;Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A;Reference number: Z21612

A;Accession: T36729

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-576 <MUR>

A;Cross-references: EMBL:AL079308; PIDN:CAB45227.1; GSPDB:GN00070; SCOEDB:SCH69.30

A;Experimental source: strain A3(2)

C;Genetics:

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C;Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 8.6%; Score 124; DB 2; Length 576;

Best Local Similarity 26.8%; Pred. No. 0.51;

Matches 74; Conservative 20; Mismatches 90; Indels 92; Gaps 19;

QY 41 SPAGTRTPPPDPFPW--VEKANTRGVGKEASKALGAKGSCETTPSRGSTPTLTPAKKN 98

DB 319 SPPATPTPTTPP-APPWQGTTPAGPSGLDRPSP---GSPGPPPTGPD---STPASPP---368

QY 99 KYRISHTPSYCDLSLFGSRSEGFAGPRMAKDAKRALLLWTPPTPRGSHSPRPRE 158

DB 369 PGTPTATGTPSAGCLPASPQDQ-----WT-PSTSGGTA--PPS 405

QY 159 APLRAIHAPGPKSTPPGPAADSOKLSMGLHSSR-PL-KRGLSHSLT---HLNVPS----209

DB 406 APSAPSAPSAPGPTRPAP-----HGTHSEVPLAERPAGMSETGSHLPPQPPTVT 456

QY 210 -TGHPTS-----APH--TNGPQDLRSTST-----GYTFERSPLVTSR 243

DB 457 PTSDAAASDAAAQAAPPAHPAPATGGPRGLUPPDRAFGRSQHPAPHPGFPPLTARS-LAPSP 515

QY 244 ARSVSI-----SVPSTPRGG-ATQKPKPP 267

DB 516 ARRADVPTAAVTARNPRSAPPQAHRGARRRRRPGPP 551

RESULT 13

T47149

Hypothetical protein DKFZp547P103.1 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 01-Dec-2000

C;Accession: T47149; E56695

R;Bloecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A;Reference number: Z24376

A;Accession: T47149

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-724 <AAA>

A;Cross-references: EMBL:AL162059

A;Experimental source: fetal brain; clone DKFZp547P103

R;Stifani, S.; Blumheller, C.M.; Redhead, N.J.; Hill, R.E.; Artavanis-Tsakonas, S.

Nature Genet. 2, 119-127, 1992

A;Title: Human homologs of a Drosophila enhancer of split gene product define a novel f

A;Reference number: A56695; MUID:93265135; PMID:1303260

A;Accession: E56695

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 277-415, 'A', 420-459, 'A', 461-724 <STI>

A;Cross-references: GB:M99439; NID:g307515; PIDN:AAA61195.1; PID:g307516

C;Genetics:

A;Gene: GDB:TL24; ESG

A;Cross-references: GDB:Z28050

A;Note: DKFZp547P103.1

C;Superfamily: unassigned WD repeat proteins; WD repeat homology

C;Keywords: nucleus

F:480-513/Domain: WD repeat homology <WD1>

F:566-599/Domain: WD repeat homology <WD2>

F:648-681/Domain: WD repeat homology <WD3>

F:689-722/Domain: WD repeat homology <WD4>

Query Match 8.6%; Score 123.5; DB 2; Length 724;

Best Local Similarity 23.2%; Pred. No. 0.69;

Matches 67; Conservative 38; Mismatches 113; Indels 71; Gaps 13;

QY 3 TPVEL-AVSGMOTLGLQRCRGYVAKRTSVVDETLFGSPAGTPTPDPPPKVVKAN 61

DB 71 TWAEINAIIGQOOLQACHLSHG-----HGLPVPLTPHFSGLOPPAIPIG 115

QY 62 RTRGVGKEASKALGAKGSC-----ETTPSRGSTPTLTPRKKNKYRPIIS 104

DB 116 SSAGL-LALSALGQSHLPIDKEKKHNDHQDRDSIKSSSVSPSASFRGAEKHR---171

QY 105 HTPSYCDLSLFGSRSEGFAGPRMAKDAKRALLLWTP---PPTPRGS--HSPR---155

DB 172 NSADYSSSESK-KQKTEKEIAARYDSGDKSDNLIIVDVSNEDPSSPRGSPAHSPRENGL 230

QY 156 -----PREAPLRAHPAGPKSTPPGPAADSOKLSMGLHSSRPLKRLGLSHSLTHLNVPS 209

DB 231 DKTRLLKKDAP---ISPASIASSSSTPSSKSELSLN-----EKSTTPVSKSN 275

QY 210 TGHPTATSA--PHTNGPQDLRSTSGVTFRSPVLTSRARSVSISVP-STP 255

DB 276 TPTPTDAPTSGNSTGRLRPVPGKPPGVDPLASSLRTPMVPCPYTP 324

RESULT 14

S29605

Q01956 rattus norv
Q14686 h nuclear r
Q91119 m nuclear r
P29353 homo sapien
O88935 mus musculu
Q9np71 homo sapien
P54282 rattus norv
P09981 rattus norv
Q9n0h3 ovis aries
Q04727 homo sapien
O88703 mus musculu
Q9y313 homo sapien

34 116 8.1 889 1 KNC3 RAT
35 116 8.1 2063 1 NCO6 HUMAN
36 116 8.1 2067 1 NCO6 MOUSE
37 115.5 8.0 583 1 SHC HUMAN
38 115.5 8.0 670 1 SYN1 MOUSE
39 115.5 8.0 852 1 WS14 HUMAN
40 115.5 8.0 2212 1 CCA1 RAT
41 115 8.0 704 1 SYN1 RAT
42 114.5 8.0 491 1 AP2A SHEEP
43 114 7.9 766 1 TLE4 HUMAN
44 114 7.9 863 1 HCN2 MOUSE
45 113.5 7.9 622 1 3Bp1_HUMAN

ALIGNMENTS

RESULT 1

SYN1_CANFA STANDARD; PRT; 415 AA.
ID SYN1_CANFA
AC O62732;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Synapsin I (Fragment).
GN SYN1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibuya H., Liu P.-C., O'Brien D.P., Johnson G.S.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
CC binds to the cytoskeleton, and is believed to function in the
CC regulation of neurotransmitter release (By similarity).
CC -!- SUBCELLULAR LOCATION: Synapse.
CC -!- SIMILARITY: Belongs to the synapsin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF049588; AAC05207.1; -.
CC HSSP; P17599; LAUX.
CC InterPro; IPR001359; Synapsin.
CC Pfam; PF02750; Synapsin_C; 1.
CC PRINTS; PR01369; SYNAPSIN.
CC PROSITE; PS00415; SYNAPSIN_1; PARTIAL.
CC PROSITE; PS00416; SYNAPSIN_2; PARTIAL.
CC Synapse; Phosphorylation; Neurone; Repeat; Actin-binding.
FT NON TER 1
FT DOMAIN <1 148 C (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT BINDING).
FT DOMAIN 149 393 D (PRO-RICH LINKER).
FT DOMAIN 384 >415 E.
FT NON TER 415
SQ SEQUENCE 415 AA; 43388 MW; E988BE2026FCS361B CRC64;

Query Match 9.7%; Score 140; DB 1; Length 415;
Best Local Similarity 26.0%; Pred. No. 0.029;
Matches 63; Conservative 17; Mismatches 108; Indels 54; Gaps 9;
Qy 40 GSPAGTRTPPPDPDPWVEKAKNRTRG-----VGKEASKALGAKGSCETTPSRGSTR 90
Db 182 GPAAQQRPP-PQGGPPQPGPGPQRPQPPQGGHLSGLGFPAG-SPLPQLPLPP 239
Qy 91 TLTPRKKNKYRPISTHTPSYCDSELSFGSRSEGSFAGPRMAKGAADAKLRALLTPPTPR- 149

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 11:58:51 ; Search time 13 Seconds
(without alignments)
1077.452 Million cell updates/sec

Title: US-10-031-589-4

Perfect score: 1437

Sequence: 1 MKTPVELAVSGMQLGLQHR.....SVPTPRRGATQKPKPMK 269

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	140	9.7	415	1 SYN1 CANFA	O62732 canis famil
2	136.5	9.5	4911	1 MLL3 HUMAN	Q8ne24 homo sapien
3	136	9.5	2716	1 OSA DROME	Q8in94 drosophila
4	134.5	9.4	503	1 WAIF HUMAN	O43516 homo sapien
5	130.5	9.1	705	1 SYN1 HUMAN	P17600 homo sapien
6	129.5	9.0	863	1 MI13 HUMAN	Q9n3f8 homo sapien
7	126.5	8.8	1199	1 P121 RAT	P52591 rattus norv
8	126.5	8.8	1321	1 IRS2 MOUSE	P81122 mus musculu
9	126	8.8	706	1 SYN1 BOVIN	P17599 bos taurus
10	125.5	8.7	699	1 VGLG HSV2H	P13290 herpes simp
11	123	8.6	886	1 VGP3 EBVA8	Q07284 Epstein-bar
12	122.5	8.5	907	1 VGP3 EBV	P33200 Epstein-bar
13	122.5	8.5	2164	1 CCA1 MOUSE	P97445 mus musculu
14	122	8.5	513	1 MI13 MOUSE	Q8bgt6 mus musculu
15	121	8.4	721	1 YK82 MYCTU	Q10690 mycobacteri
16	120.5	8.4	753	1 ELA2 HUMAN	Q8iyf1 homo sapien
17	120.5	8.4	1083	1 T2D3 HUMAN	O00268 homo sapien
18	120.5	8.4	1125	1 IF22 FROM	Q7v544 prochloroco
19	120.5	8.4	2142	1 BAT2 HUMAN	P48634 homo sapien
20	120	8.4	779	1 NIM1 NEUCR	P48479 neurospora
21	120	8.4	870	1 SRC2 HUMAN	Q86gp6 homo sapien
22	120	8.4	3530	1 MY15 HUMAN	Q9ukn7 homo sapien
23	119	8.3	1446	1 I818 PRVKA	P33479 pseudorabie
24	118.5	8.2	677	1 OQFR HUMAN	Q9nzt2 homo sapien
25	118.5	8.2	1077	1 HLES DROME	Q92308 drosophila
26	117.5	8.2	757	1 KNC3 HUMAN	Q40003 homo sapien
27	117	8.1	1207	1 AT19 HUMAN	Q8te59 homo sapien
28	117	8.1	1461	1 IE18 PRVIF	P1675 pseudorabie
29	116.5	8.1	1198	1 HCN4 RAT	Q9jka7 rattus norv
30	116.5	8.1	3511	1 MY15 MOUSE	Q9gz24 mus musculu
31	116	8.1	530	1 NUPA MOUSE	Q8bg30 mus musculu
32	116	8.1	681	1 PAK6 HUMAN	Q9nq45 homo sapien
33	116	8.1	766	1 TLE4 MOUSE	Q62441 mus musculu

Db 240 TSAPO-----QVSAQPLSQAGQSPVAGGFGAPPAAPSPSPQAGFPQATQ 294
 QY 150 -----GSHSPREAPLRAHPAGSKTEPGPAASQKLSMGLHSSRLKGLSHSLTHL 205
 Db 295 TSVSGQAPFKASGAPSPQKQKPPGAGPAGTRQASQAG----- 336
 QY 206 NVPSTGHATAPHTNGPQDLRPSTSGVTFPSPLVTSRARSVSVSTPRRGATOKPK 265
 Db 337 PMPTGPTTQPP-----RPSGPGAGR-PAKPOLAQKPSQDVSPATAG----- 381
 QY 266 PP 267
 Db 382 PP 383

RESULT 2
 MLL3 HUMAN
 ID MLL3 HUMAN STANDARD; PRT; 4911 AA.
 AC Q8NEZ4; Q8NC02; Q8NDF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-
 DE lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43)
 DE (Homologous to ALR protein).
 GN MLL3 OR HALR OR KIAA1506.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fetal thymus;
 RX MEDLINE=21888622; PubMed=11891048;
 RA Rualt M., Brun M.-E., Ventura M., Roizes G., De Sario A.;
 RT "MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a
 RT chromosome region frequently deleted in myeloid leukaemia.";
 RL Gene 284:73-81(2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=21574953; PubMed=11718452;
 RA Tan Y.C., Chow V.T.;
 RT "Novel human HALR (MLL3) gene encodes a protein homologous to ALR and
 RT to ALI-1 involved in leukemia, and maps to chromosome 7q36 associated
 RT with leukemia and developmental defects.";
 RL Cancer Detect. Prev. 25:454-469(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22737999; PubMed=12853948;
 RA Hiller L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
 RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
 RA Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
 RA Fewell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
 RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
 RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
 RA Orzsky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
 RA Strong C.W., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
 RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
 RA Tin-Wollam A.-M., Abbott A., Minx P., Maupin R., Strommatt C.,
 RA Latreille P., Miller N., Johnson D., Murray J., Woessner J.P.,
 RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
 RA Bieri T.A., Nelson J.O., Berkowicz N., Wohlmann P.E., Cook L.L.,
 RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
 RA Clifton S.W., Chisoe S.B., Marra M.A., Raymond C., Haugen E.,
 RA Gillett W., Zhou Y., James R., Phelps K., Iadonato S., Bubb K.,
 RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
 RA Baertsch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
 RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
 RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
 RA Waterston R.H., Wilson R.K.;
 RT "The DNA sequence of human chromosome 7.";

RL Nature 424:157-164(2003).
 RN [4]
 RP SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=20277482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 RT the complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 RN [5]
 RP SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Nimomiya K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 3879-4911 FROM N.A.
 RC TISSUE=Testis;
 RA Duesterhoef A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=22371496; PubMed=12482968;
 RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
 RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
 RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.-J., Lee K.-J., Lee Y.C.,
 RA Lee J.W.;
 RT "Activating signal cointegrator 2 belongs to a novel steady-state
 RT complex that contains a subset of trithorax group proteins.";
 RL Mol. Cell. Biol. 23:140-149(2003).
 CC [1-] FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
 CC coactivator complex of nuclear receptors, involved in
 CC transcriptional coactivation. MLL3 may be a catalytic subunit of
 CC this complex, which weakly methylates Lys-4 of histone H3. This is
 CC a specific tag for epigenetic transcriptional activation. May be
 CC involved in leukemogenesis and developmental disorder.
 CC [2-] CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
 CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC [3-] SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
 CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
 CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
 CC MLL2 and MLL3, and ASH2/ASCL2. Interacts with histone H3.
 CC [4-] SUBCELLULAR LOCATION: Nuclear (Probable).
 CC [5-] ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8NEZ4-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8NEZ4-2; Sequence=VSP_008561, VSP_008562;
 CC [6-] TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed
 CC by brain and liver. Also expressed in placenta, peripheral
 CC blood, fetal thymus, heart, lung and kidney. Within brain,
 CC expression was highest in hippocampus, caudate nucleus, and
 CC substantia nigra. Not detected in skeletal muscle and fetal liver.
 CC [7-] DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
 CC and H4, and may have a H3 lysine specific methylation activity.
 CC [8-] MISCELLANEOUS: Found in a critical region of chromosome 7, which
 CC is commonly deleted in malignant myeloid disorders. Partial
 CC duplication of the MLL3 gene are found in the juxtacentromeric
 CC region of chromosomes 1, 2, 13 and 21. Juxtacentromeric
 CC reshuffling of the MLL3 gene has generated the BAGE genes.
 CC [9-] SIMILARITY: Belongs to the TRX/MLL family.
 CC [10-] SIMILARITY: Contains 1 DHHC-type zinc finger.
 CC [11-] SIMILARITY: Contains 6 PHD-type zinc fingers.
 CC [12-] SIMILARITY: Contains 1 post-SET domain.

CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF024361; AK00583.1; -;
CC EMBL; AF264750; AAF74766.2; -;
CC EMBL; AC006017; AAD45822.1; -;
CC EMBL; AC104692; -; NOT ANNOTATED CDS.
CC EMBL; AC005631; -; NOT ANNOTATED_CDS.
CC EMBL; AB040939; BAA96030.2; -;
CC EMBL; AK022687; BAB14179.1; -;
CC EMBL; AK075113; BAC11409.1; -;
CC EMBL; AL833924; CAD38780.1; -;
CC Genew; HGNC:13726; MLL3.
CC MIM; 606833; -;
CC InterPro; IPR000637; AT hook.
CC InterPro; IPR003889; FYrich_C.
CC InterPro; IPR003888; FYrich_N.
CC InterPro; IPR000910; HMG_12_box.
CC InterPro; IPR003616; PostSET.
CC InterPro; IPR001214; SET.
CC InterPro; IPR001594; Znf_DHHC.
CC InterPro; IPR001965; Znf_PHD.
CC InterPro; IPR001841; Znf_Ring.
CC Pfam; PF00505; HMG_box; 1.
CC Pfam; PF00628; PHD; 6.
CC Pfam; PF00856; SET; 1.
CC SMART; SM00542; FYR; 1.
CC SMART; SM00541; FYRN; 1.
CC SMART; SM00398; HMG; 1.
CC SMART; SM00249; PHD; 8.
CC SMART; SM00508; PostSET; 1.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS00354; HMG1_Y; 1.
CC PROSITE; PS00868; POST_SET; 1.
CC PROSITE; PS00280; SET; 1.
CC PROSITE; PS0216; ZF_DHHC; 1.
CC PROSITE; PS01359; ZF_PHD; 1; 5.
CC PROSITE; PS00016; ZF_PHD; 2; 6.
CC PROSITE; PS00089; ZF_RING; 2; 1.
CC Transferrase; Methyltransferase; Chromatin regulator; Activator;
CC DNA-binding; Nuclear protein; Transcription regulation; Coiled coil;
CC Zinc-finger; Repeat; Alternative splicing; Polymorphism.
CC ZN_FING; 341 391 PHD-TYPE 1.
CC ZN_FING; 344 389 RING-TYPE.
CC ZN_FING; 388 438 PHD-TYPE 2.
CC ZN_FING; 436 489 DHHC-TYPE.
CC ZN_FING; 464 520 PHD-TYPE 3.
CC ZN_FING; 957 1010 PHD-TYPE 4.
CC ZN_FING; 1007 1057 PHD-TYPE 5.
CC ZN_FING; 1084 1139 PHD-TYPE 6.
CC ZN_FING; 1470 4891 SET.
CC DOMAIN; 4770 4891 POST-SET.
CC DOMAIN; 4895 4911 COILED COIL (POTENTIAL).
CC DOMAIN; 92 112 COILED COIL (POTENTIAL).
CC DOMAIN; 644 672 COILED COIL (POTENTIAL).
CC DOMAIN; 1338 1366 COILED COIL (POTENTIAL).
CC DOMAIN; 1754 1787 COILED COIL (POTENTIAL).
CC DOMAIN; 3054 3081 COILED COIL (POTENTIAL).
CC DOMAIN; 3272 3433 COILED COIL (POTENTIAL).
CC DOMAIN; 3391 3433 COILED COIL (BY SIMILARITY).
CC DOMAIN; 34 46 A.T HOOK (BY SIMILARITY).
CC DOMAIN; 1719 1796 GLN-RICH.
CC DOMAIN; 1834 2281 PRO-RICH.
CC DOMAIN; 2412 2630 PRO-RICH.
CC DOMAIN; 2690 2786 ASP-RICH.

Query Match 9.58; Score 136.5; DB 1; Length 4911;
Best Local Similarity 24.08; Pred. NO. 0.57;
Matches 75; Conservative 28; Mismatches 113; Indels 97; Gaps 13;
QY 29 ARTSYVDTEFLGSPAGTR-----PTPPDFDPFWVK-----AMTRGVGK 68
DB 1933 SRPLQWNETTANRSPVYDLCSSTTNDVYAKPDPFVMTDQFSLGLSRSPVUSE 1992
QY 69 EASKALGAKGSC-----ETTPSRGSGTPTLTPRKVK-----YRPISTPS 108
DB 1993 QTAGPIAAGTSDHFTKPSRADYVQRIQIPDSYARELLTPAFLDSGSGFPTQMPPPS 2052
QY 109 YCDESLFGSRSEGAFCAPMAKGAALRALWTPPTPRGSH--SPRPREAPLRAIHP 166
DB 2053 SQDP--YGSVSG-----ASRLSDVPIERPAL--TPRPIDNESHNQSDPYSQLPLPHP 2103
QY 167 -----AGPSK--TEPG-----PAADSOXLKSMGLHS-----SRPL 194
DB 2104 AVNESFAHPSRAFSQPGTISRPTSDQPSYQPPGPRPRWDSYSQSSGTAASNTDPSYQP- 2162
QY 195 KRGLSHSLTHLNVSTGHPATSAHTNGPDRLRSTSGVTFRSLVTSRARSVSISVPST 254
DB 2163 -----FGTPRPTVDPSYQQPQTPRPRSTQTDLFVTPVTNQRHSDPYAHPPGT 2209
QY 255 PRRGATOKPKPP 267
DB 2210 PRGISVPSYQPP 2222
RESULT 3
OSA_DROME STANDARD; PRT; 2716 AA.
ID OSA_DROME STANDARD; PRT; 2716 AA.
AC Q8TN94; O61603; Q9VEG7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trithorax group protein OSA (Eyelid protein).
GN OSA OR ELD OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=97415319; PubMed=9271118;
RX Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
RT and has homology to the Bright family of DNA-binding proteins."
RL Genes Dev. 11:1949-1962(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

QY 68 KEASKALGAKGSCETTPSRGSTPTLTTPRKKNKYPDISHTPSYCDSESLFSGRSEGASFGAP 127
Db 689 GANSMFSGAQAG--GYFGRG-----MFINHTGOYPPQWVPPSPQQTVPQGAFFGAMVGNH 741
QY 128 RMAKGAAPKALLWTP-----PPTPRGSHSPRPREAPLRAIHAPGSKTEPFGPAADS 180
Db 742 VQKGG-----TPPPFVGVGGPPPPQGGSPREFNLVYKQHLQHKGGYGGSPFPFQGP 791
QY 181 QKLSMG--GLHSSRPV-----KRLGSLHSLHNVPT-----GHPATSAPHT 220
Db 792 QYGVNGPTGHPMPMGPHNGPPHGPPTNMGPPTSTPPQSQVLOGGQPGGQASGGPES 851
QY 221 NGQDLRPSTGVTFRSPLVTSRARSVSISVPSTP 255
Db 852 GGPEHI--SDNGISSSGPTGAAGNHAVTSVVTGP 885

RESULT 4
WASP HUMAN STANDARD; PRT; 503 AA.
ID: WAP1 HUMAN O43516; Q15220; Q9UNP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Wiskott-Aldrich syndrome protein interacting protein (WASP) interacting
DE protein) (PRL-2 protein).
OS WASP OR WIP.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98070810; PubMed=9405671;
RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT induces actin polymerization and redistribution in lymphoid cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
[2]
RN SEQUENCE FROM N.A.
RX TISSUE=Tonsil;
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX MEDLINE=99218549; PubMed=10202051;
RA Stewart D.M., Tian L., Nelson D.L.;
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT interacting protein.";
RL J. Immunol. 162:5019-5024(1999).
CC -!- FUNCTION: May have direct activity on the actin cytoskeleton.
CC Induces actin polymerization and redistribution.
CC -!- SUBUNIT: Binds to WASP, profilin and actin.
CC -!- TISSUE SPECIFICITY: Highly expressed in peripheral blood
CC mononuclear cells, spleen, placenta, small intestine, colon,
CC thymus. Lower expression in ovary, heart, brain, lung, liver,
CC skeletal muscle, kidney, pancreas, prostate and testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF031588; AAC03767.1; -;
CC EMBL; AF106062; AAD45972.1; -;
CC EMBL; X86019; CAA60014.1; -;
CC Genew; HGNC:12736; WASP.
CC MIM; 602357; -;

DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0003779; F:actin binding; TAS.
DR GO; GO:0005522; F:profilin binding; TAS.
DR GO; GO:0008154; P:actin polymerization and/or depolymerization; TAS.
DR GO; GO:0006461; P:protein complex assembly; TAS.
DR InterPro: IPR003124; WH2.
DR Pfam: PF02205; WH2; 1.
DR SMART: SMO0246; WH2; 1.
KW Actin-binding; Repeat; Polymorphism.
FT DOMAIN 2 13 POLY-PRO.
FT DOMAIN 64 96 GLY-RICH.
FT DOMAIN 241 244 POLY-SER.
FT DOMAIN 362 433 PRO-RICH.
FT REPEAT 352 361 XRSQGXPPXP MOTIF 1.
FT REPEAT 374 383 XRSQGXPPXP MOTIF 2.
FT REPEAT 410 419 XRSQGXPPXP MOTIF 3.
FT SITE 45 48 ACTIN BINDING.
FT VARIANT 495 495 G -> A.
FT /FTID=VAR_010295.
FT CONFLICT 303 309 PHRPLR -> SSOAPP (IN REF. 3).
FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRERGPPPLPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;
Query Match 9.4%; Score 134.5; DB 1; Length 503;
Best Local Similarity 25.8%; Pred. No. 0.077;
Matches 63; Conservative 23; Mismatches 107; Indels 51; Gaps 11;
QY 42 PAGTR-PTPPDFDPWVEKANTRGVGKEASKALGAKGSCETTPSRGSTPT--LTPRKN 98
Db 207 PGGPQPSPGPTPPPPFGNRTALGGGSIROSLSSSPFNRRPLPPTPSRALDDKPP 266
QY 99 KYRPISTPSYCDSESLFSGRSEGASFGAPRMAKGAALRALLWTPPTPRGSHSPRPR- 157
Db 267 PPPVGNRPSHRAV-----PPPPQNNKPPVPST 297
QY 158 ---EAPLRA-IHPAGPSTKTEFGPAADSOGLSMGLHSSRPLKRLGSLHSLHNVPTSGH- 212
Db 298 PPSAPHPHRLPPPPSRPPGPPPLPS---SSGNDETPLRQLNLSLSSTPTPLPSPGRS 354
QY 213 -PATSAPHTNGPQDLR--PSTSGVTFRSPLVTSRARSVSISVPSTP---RGGATQK--- 263
Db 355 GLPPLPPPSERPPPPVDRDPGRSG-PLPPPPVVRNGSTSRALPATPQLPSRSGVDSRSG 413
QY 264 PKPP 267
Db 414 PRPP 417
RESULT 5
ID: SYN1 HUMAN STANDARD; PRT; 705 AA.
AC P17600; O75825;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Synapsin I (Brain protein 4.1).
GN SYN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90243651; PubMed=2110562;
RA Suedhof T.C.;
RT "The structure of the human synapsin I gene and protein.";
RL J. Biol. Chem. 265:7849-7852(1990).
[2]
RN SEQUENCE OF 1-125 FROM N.A.
RX MEDLINE=90368667; PubMed=2118519;
RA Sauerwald A., Hoesche C., Oschwald R., Killmann M.W.;

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dami P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming L., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tanlyn G., Heathcott R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mccloy J., McLaren S., McMuray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuoyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Descamps S., Do A., Do T.,
RA Dorman A., Fang F., Hu F., Hua A., Keston S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull B., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bentley G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaik T., Kurahashi H., Saitta B.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edlman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.,
RA "The DNA sequence of human chromosome 22";
RA Nature 402:489-495(1999).
[3]
SEQUENCE OF 25-863 FROM N.A.
RA TISSUE=Melanoma;
RA Ansoxge W., Winkner U., Mewes H.-W., Weil B., Wiemann S.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE OF 73-863 FROM N.A.
RA TISSUE=Brain;
RA MEDLINE=21156230; PubMed=11258795;
RA Hirotsawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
RA "Identification of novel transcribed sequences on human chromosome 22
by expressed sequence tag mapping";
RA DNA Res. 8:1-9(2001).
[5]
SEQUENCE OF 514-863 FROM N.A.
RA TISSUE=Choriocarcinoma;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Stapleton M., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalls D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";

RESULT 7
P121_RAT

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC - SUBUNIT: Binds to Rab13.
CC - SIMILARITY: Contains 1 calponin-homology (CH) domain.
CC - SIMILARITY: Contains 1 LIM zinc-binding domain.
CC - CAUTION: Ref.3 (CAD39036) sequence differs from that shown due to
frameshifts in positions 486 and 507.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: AJ496196; CAD42713.1; -
EMBL: AL022311; CAB62980.1; -
EMBL: AL022311; CAB62979.1; -
EMBL: AL022311; CAB62978.1; ALT_SEQ.
EMBL: AL834373; CAD39036.1; ALT_FRAME.
EMBL: AL833860; CAD38718.1; -
EMBL: AB051455; BAB33338.1; -
EMBL: BC001090; AAH01090.2; -
HSSP: Q01082; 1BKE.
GO: GO:0005622; C:intracellular; IC.
GO: GO:0016192; P:vesicle-mediated transport; IC.
InterPro: IPR001715; Calponin-like.
InterPro: IPR001781; LIM.
Pfam: PF00307; CH; 1.
Pfam: PF00412; LIM; 1.
ProDom: PD000094; LIM; 1.
SMART: SM00033; CH; 1.
SMART: SM00132; LIM; 1.
PROSITE: PS00021; CH; 1.
PROSITE: PS00478; LIM DOMAIN 1; 1.
PROSITE: PS00023; LIM DOMAIN 2; 1.
LIM domain, Metal-binding; Zinc.
FT DOMAIN 2 105
FT DOMAIN 162 225
FT DOMAIN 253 667
FT DOMAIN 428 437
FT CONFLICT 137 137 P -> S (IN REF. 4).
FT CONFLICT 212 212 F -> Y (IN REF. 3); CAD39036).
FT CONFLICT 519 519 A -> S (IN REF. 3).
SEQUENCE 863 AA; 93441 MW; 755E3B57C6037292 CRC64;
Query Match 9.0%; Score 129.5; DB 1; Length 863;
Best Local Similarity 25.9%; Pred. No. 0.27;
Matches 64; Conservative 30; Mismatches 130; Indels 23; Gaps 11;
QY 40 GSPAGTRTP-PDFPPW---VEKANRTGVGKASKALGAKGSCETTPSRGST----P 90
Db 354 GTPKPSGTPAPRKPPWITLVQAEPKKPPAPLPSSSGPPSPQSRQVENGTEVAQP 413
QY 91 TLTPKKNK-YRPIGHTSPSYCDSEIFGSRSEGAFGAFMAKGAALRLW---TPPP 146
Db 414 STASLESKPYNPFEEEDKEEAPAPSLATS---PALGHPESTPKSLHPWYGTTPS 470
QY 147 TTRGSHSPRPR---EAPLAIHPAGPKTEPGPADSQKLSMGLHSSRPKKGLSHSLT 203
Db 471 SPRTKKRPAPRAPSPASPL-ALFASRLSHSEPPSATPSPALSVESLSSESASATAGAELE 529
QY 204 HLNVP-STGHPTATSPHTNGPQDLRPSVSGVTFRSPLVTSRARSVSISVPS-TPR-RGGA 260
Db 530 PPAVPKSSSEPAVHAGTGNPNVSLTNSLASSGELVEPRVEQMPQASPLAPTRGSS 589
QY 261 TQPKPP 267
Db 590 GQPAKE 596


```

882 PTRLSLEGLTLPQMDEYPTPEPKSPGEVIN--IDPGEAGTSLSP--APPLISAASS 937
64 RGVGKAKA--LCAKGCSTTSPRGSTPTLTTRKKNKVPISHTSYCDESLFGSR--- 118
938 SSLISASSPSSGLSGTPTGTSRQSPSLDYMNLDPSPKSPKPTSTRSGDTVGSMDGL 997
119 -SEGAGPAPMAKGAALRALLWTPTTPRGSHSPRPREAPLRAIHPAGPS---KTEP 174
998 LSPFAS--SP-----YPLPFRSTSPSSLOQL-----PPAPGDLVRLPP 1036
175 GPAADSQLSMGLHSGRLPKRGLSHSLTHLNVST-GHPATSAHTNGPQDLRPSTGV 233
1037 ASAATSCQTAGSMSEPGDNG-DYSEMAFGVAATPPQIVAPPKPEGARVASP-TSG 1094
234 TFRSPL--VTSRARSVISVPSPRRGATQKPKP 266
1095 KRUSLMDQVSGVEAFICVSPDPHKGAKVIRADP 1129

RESULT 9
SYNL_BOVIN
ID -SYNL_BOVIN STANDARD; PRT; 706 AA.
AC P17599;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Synapsin I.
GN SYNL
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=89388265; PubMed=2506642;
RA Suedhof T.C., Czernik A.J., Kao H.-T., Takei K., Johnston P.A.,
RA Horiuchi A., Kanazir S.D., Wagner M.A., Perin M.S., de Camilli P.,
RA Greengard P.;
RT "Synapsins: mosaics of shared and individual domains in a family of
RT synaptic vesicle phosphoproteins.";
RL Science 245:1474-1480(1989).
[2]
RN [3]
RP PHOSPHORYLATION SITE BY PROLINE-DIRECTED KINASE.
RX MEDLINE=90216728; PubMed=2108963;
RA Hall F.L., Mitchell J.P., Vulliamt P.R.;
RT "Phosphorylation of synapsin I at a novel site by proline-directed
RT protein kinase.";
RL J. Biol. Chem. 265:6944-6948(1990).
[3]
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 112-417.
RX MEDLINE=98130589; PubMed=9463376;
RA Esser L., Wang C.R., Hosaka M., Smagula C.S., Suedhof T.C.,
RA Deisenhofer J.;
RT "Synapsin I is structurally similar to ATP-utilizing enzymes.";
RL EMBO J. 17:977-984(1998).
CC -!- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
CC binds to the cytoskeleton, and is believed to function in the
CC regulation of neurotransmitter release.
CC -!- SUBCELLULAR LOCATION: Synapse.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Synapsin IA;
CC IsoId=P17599-1; Sequence=Displayed;
CC Name=Synapsin IB;
CC IsoId=P17599-2; Sequence=VSP_006314, VSP_006315;
CC -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -!- SIMILARITY: Belongs to the synapsin family.
-----
```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27810; AAA30761.1; -
DR EMBL; M27811; AAA30762.1; -
DR PDB; 1AUV; 18-MAR-98.
DR PDB; 1AUX; 18-MAR-98.
DR GlycoSuiteDB; P17599; -
DR InterPro; IPR001359; Synapsin.
DR Pfam; PF02078; Synapsin_1.
DR PRINTS; PR01368; SYNAPSIN.
DR PROSITE; PS00415; SYNAPSIN_1; 1.
DR PROSITE; PS00416; SYNAPSIN_2; 1.
KW Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW Alternative splicing; 3D-structure.
FT DOMAIN 1 28 A.
FT DOMAIN 29 112 B (LINKER).
FT DOMAIN 113 420 C
FT (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT BINDING).
FT D (PRO-RICH LINKER).
FT E.
FT PHOSPHORYLATION (BY PKA AND CAMK1).
FT PHOSPHORYLATION (BY PDPK).
FT MOD_RES 551 551 PHOSPHORYLATION (BY CAMK2).
FT MOD_RES 568 568 PHOSPHORYLATION (BY CAMK2).
FT MOD_RES 605 605 PHOSPHORYLATION (BY CAMK2).
FT VARSPLOC 662 670 NKSQLTNA -> KASPAQAQP (in isoform
FT Synapsin IB).
FT /FTID=VSP_006314.
FT Missing (in isoform Synapsin IB).
FT /FTID=VSP_006315.
FT
FT VARSPLIC 671 706
FT
FT STRAND 114 119
FT TURN 122 123
FT HELIX 126 130
FT TURN 131 132
FT STRAND 134 135
FT TURN 136 138
FT STRAND 139 146
FT TURN 149 150
FT STRAND 151 155
FT TURN 157 158
FT STRAND 161 165
FT TURN 172 173
FT STRAND 176 178
FT STRAND 182 185
FT STRAND 190 190
FT TURN 194 195
FT STRAND 197 197
FT HELIX 199 207
FT TURN 208 209
FT STRAND 212 213
FT HELIX 216 221
FT TURN 222 223
FT HELIX 225 229
FT TURN 241 243
FT STRAND 246 246
FT STRAND 250 252
FT HELIX 255 257
FT STRAND 266 270
FT TURN 275 276
FT STRAND 279 282
FT HELIX 285 296
FT TURN 297 299
FT STRAND 302 306
FT STRAND 310 319
FT TURN 320 321
FT STRAND 322 327
```

```

FT STRAND 345 347
FT HELIX 351 360
FT TURN 361 362
FT HELIX 363 365
FT STRAND 369 377
FT TURN 378 379
FT STRAND 382 388
FT TURN 390 391
FT HELIX 399 415
SQ SEQUENCE 706 AA; 74518 MW; 89373750BF014340 CRC64;

Query Match 8.8%; Score 126; DB 1; Length 706;
Best Local Similarity 22.9%; Pred. No. 0.37;
Matches 60; Conservative 23; Mismatches 87; Indels 92; Gaps 12;

QY 40 GSPAGTRTPDFDPWVEKANTRG-----VKEAKALGAKGSCETTPSRGSTP 90
DB 454 GPFAQCRP-PFGGPPQPGPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRPQRP 511
QY 91 TLTRKKNKYPISHTPSYCDSELSFGSRSEGSFGAPRMAKDAKRLALLWTP--PPTP 148
DB 512 TSVPQ-----QPASQ-----ATPMTQGGQGRQSRPVAGFGAPPAT 546
QY 149 RGSHPRR-----EAPLR--IHPA-----GPSKTEPGAADSOQLSM 185
DB 547 RPPASPSQROAGPPOATQTSVSCQAPPKASGVPPGQQRQRPQRPQRPQRPQRPQ 606
QY 186 GGLHSSRLKRLSHLSLHNLNVPSTGHPTSAHTNGPQDLRPTSTSGVTFERSPLVTSR 245
DB 607 AG-----PMPTGPTTQQRPSGGPAGRTKPLQAKP----- 641
QY 246 SVSISVSTPRRGATQKPKPP 267
DB 642 SQDVPPDATAAGG-----PP 657

RESULT 10
VGLG_HSV2H STANDARD; PRT; 699 AA.
AC P13290;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Glycoprotein G.
GN GG OR US4.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8711457; PubMed=3027242;
RA McGeoch D.J., Moss H.W.M., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the
RT short unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons";
RL J. Gen. Virol. 68:19-38(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC 2: GH, GB, GC, GD, GI, AND GE.
CC -1- MISCELLANEOUS: GLYCOPROTEIN G IS MUCH LARGER IN HSV-2 THAN IN
CC HSV-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC
DR EMBL; X04798; CAA28490.1; -
DR EMBL; Z86099; CAB06711.1; -
DR PIR; C43674; C43674.
DR InterPro: IPR003363; Herpes GG.
DR InterPro: IPR007110; Ig-like.
DR Pfam; PF02400; Herpes GG; 1.
KW Glycoprotein.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 699 AA; 72243 MW; 612BA7B345E75540 CRC64;

Query Match 8.7%; Score 125.5; DB 1; Length 699;
Best Local Similarity 23.0%; Pred. No. 0.39;
Matches 59; Conservative 27; Mismatches 92; Indels 79; Gaps 11;

QY 35 DETLFGSPAGTRTPDFDPWVEKANTRGUGKEAKALGAKGSCETTPSRGSTPTLTP 94
DB 405 EETAVASPPATASVE---SSLPFAAAAATPGAGHTNTSSASAAKTPPTTPA---PTTP 457
QY 95 RKKNKYRPISTPSYCDSELSFGSRSEGSFGAPRMAKDAKRLALLWTPP---PTPRG 150
DB 458 PTHATPRTPP-----GPQT-----TPGPATPGPVG 486
QY 151 SHSPRPRAPLRAHPAGPSTKTEPGAADSOQLSMGGLHSSRLKRLGLSHS-----LTHL 205
DB 487 ASAAPATDSPLTASPPA---TAPGSAANVSVA---ATTATPGTGTARTPTDPKTHP 539
QY 206 NVPSTGHEATSAPE---HTNGPQDLR-----PSTSGVTFRSPLVTSRARSVSIS 250
DB 540 HCPADAPGSPAPPPPHRGGEFEFGEGDGEDDDSDATGLAFRTPNPK----- 591
QY 251 VPSTPRRGATQKPKPP 267
DB 592 --PPARPGRPTLTP 606

RESULT 11
VGP3_EBVA8 STANDARD; PRT; 886 AA.
AC Q07284;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Envelope glycoprotein GP340 (Membrane antigen) (NA).
GN BLUFI.
OS Epstein-barr virus (strain AG876) (Human herpesvirus 4), and
OS Epstein-barr virus (strain P3HR-1) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=82830, 82829;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AG876, and P3HR-1;
RX MEDLINE=93331716; PubMed=8393237;
RA Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,
RA Arrand J.R.;
RT "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
RT conserved between virus types A and B.";
RL Virology 195:578-586(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P3HR-1;
RA Klein K., Mueller-Lantzsch N.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC B-CELLS.
CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC ENVELOPE.

```


CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; L07922; AAA02783.1; -			CC	EMBL; V01555; CAA24854.1; -		
DR	EMBL; L07923; AAA02787.1; -			DR	PIR; A43042; QQB21.		
DR	EMBL; X67776; CAA47986.1; -			DR	PIR; B43042; QQB22.		
DR	PIR; S29605; S29605.			DR	PIR; P05109; Herpes_BULF1.		
DR	InterPro; IPR007796; Herpes_BULF1.			DR	InterPro; IPR007796; Herpes_BULF1.		
DR	Membrane; Glycoprotein; Antigen; Late protein.			DR	Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.		
KW	CARBOHYD 47			KW	CARBOHYD 47		
FT	CARBOHYD 87			FT	CARBOHYD 87		
FT	CARBOHYD 114			FT	CARBOHYD 114		
FT	CARBOHYD 166			FT	CARBOHYD 166		
FT	CARBOHYD 169			FT	CARBOHYD 169		
FT	CARBOHYD 195			FT	CARBOHYD 195		
FT	CARBOHYD 229			FT	CARBOHYD 229		
FT	CARBOHYD 277			FT	CARBOHYD 277		
FT	CARBOHYD 318			FT	CARBOHYD 318		
FT	CARBOHYD 328			FT	CARBOHYD 328		
FT	CARBOHYD 345			FT	CARBOHYD 345		
FT	CARBOHYD 356			FT	CARBOHYD 356		
FT	CARBOHYD 378			FT	CARBOHYD 378		
FT	CARBOHYD 386			FT	CARBOHYD 386		
FT	CARBOHYD 411			FT	CARBOHYD 411		
FT	CARBOHYD 435			FT	CARBOHYD 435		
FT	CARBOHYD 443			FT	CARBOHYD 443		
FT	CARBOHYD 457			FT	CARBOHYD 457		
FT	CARBOHYD 497			FT	CARBOHYD 497		
FT	CARBOHYD 519			FT	CARBOHYD 519		
FT	CARBOHYD 533			FT	CARBOHYD 533		
FT	CARBOHYD 568			FT	CARBOHYD 568		
FT	CARBOHYD 589			FT	CARBOHYD 589		
FT	CARBOHYD 624			FT	CARBOHYD 624		
FT	CARBOHYD 680			FT	CARBOHYD 627		
FT	CARBOHYD 886 AA; 92388 MW; 4394F6130DCC8A CRC64;			FT	CARBOHYD 645		
SQ	SEQUENCE			FT	CARBOHYD 656		
Query Match				EMBL; V01555; CAA24854.1; -			
Best Local Similarity 23.9%; Pred. No. 0.72;				PIR; A43042; QQB21.			
Matches 79; Conservative 30; Mismatches 120; Indels 102; Gaps 15;				PIR; B43042; QQB22.			
Qy	3	TPVELAVSGMQLGLQHCRCRGVYKARTSVVDETLFGSPAGTRPTP-----PDFDPPW 56		Qy	3	TPVELAVSGMQLGLQHCRCRGVYKARTSVVDETLFGSPAGTRPTP-----PDFDPPW 56	
Db	529	TPTPNATP-----PTLG-----KTSFTSVITPTFNATSPVAVTPTPTNATPT 573		Db	529	TPTPNATP-----PTLG-----KTSFTSVITPTFNATSPVAVTPTPTNATPT 573	
Qy	57	VEKANRTG-----VGKEASKA-----LGAKGSCB--TTPSRGSGTPTLTPRK 96		Qy	57	VEKANRTG-----VGKEASKA-----LGAKGSCB--TTPSRGSGTPTLTPRK 96	
Db	574	LKTSPTSVAVTPTPTNATSPVGTGSPQANTNHTLGTSTSTPVVTSPPKNATSAVTTGQ 633		Db	574	LKTSPTSVAVTPTPTNATSPVGTGSPQANTNHTLGTSTSTPVVTSPPKNATSAVTTGQ 633	
Qy	97	KN-----KYRPISTHPSYCDSELFGRSEGASFGAPMAKG-----DAAKLRAL 140		Qy	97	KN-----KYRPISTHPSYCDSELFGRSEGASFGAPMAKG-----DAAKLRAL 140	
Db	634	HNITSSSTSMRLSPSSISELSTSTSDNSTSHMPLTSAHPGGENITQVTPASTSTHH 693		Db	634	HNITSSSTSMRLSPSSISELSTSTSDNSTSHMPLTSAHPGGENITQVTPASTSTHH 693	
Qy	141	LWTPPTPTPR-----GSHSRPRPREA-----PLRAHPAGES-----KTE 173		Qy	141	LWTPPTPTPR-----GSHSRPRPREA-----PLRAHPAGES-----KTE 173	
Db	694	VSTSSPAPRPTTQASGPGNSSTSTKPGEVNVTGPKPNATSPQAPSGQKTAAPTPTS 753		Db	694	VSTSSPAPRPTTQASGPGNSSTSTKPGEVNVTGPKPNATSPQAPSGQKTAAPTPTS 753	
Qy	174	PGPAADSKLGMGLHHSRPLKRLGSLHSLHNVPTST---GHPATSAFHTNGQDLRPS 230		Qy	174	PGPAADSKLGMGLHHSRPLKRLGSLHSLHNVPTST---GHPATSAFHTNGQDLRPS 230	
Db	754	TGGKANS---TTGGKHTT-----GHGARTSTSTFTTDYGGDSTTPTRYNATLYLPPT 803		Db	754	TGGKANS---TTGGKHTT-----GHGARTSTSTFTTDYGGDSTTPTRYNATLYLPPT 803	
Qy	231	SGV-----TPRSPLVTSRARSVSVSPSTPR 256		Qy	231	SGV-----TPRSPLVTSRARSVSVSPSTPR 256	
Db	804	SSKLAPRWTFISPPVTTAQAATVPVPPISQPR 834		Db	804	SSKLAPRWTFISPPVTTAQAATVPVPPISQPR 834	
RESULT 12				RESULT 12			
VGP3_EBV				VGP3_EBV			

FT CARBOHYD 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 858 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 888 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 502 Missing (in isoform GP220).
 FT FTid=VSP_002070.
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBA5C529 CRC64;
 Query Match 8.5%; Score 122.5; DB 1; Length 907;
 Best Local Similarity 22.8%; Pred. No. 0.79;
 Matches 77; Conservative 24; Mismatches 128; Indels 109; Gaps 13;
 QY 3 TPVELAVSGWOTGLQHCRCRGYRVKARTSVYDTLFGSPAGTRPT-----P 49
 Db 543 TPTNAT--PTLG-----KTSPTSAVTTPTPNATSPTLGKTSPTS AVTTPT 587
 QY 50 PDFDPVWEKANRTRGVGKEASKALG-----AKGSCETTPSRGSTTLTPRKKNKY 100
 Db 588 PNATSPILGKTSPTS AVTTPTPNATGTGVTGTSPOANATNHTLGGTSPTPVTSQPKNAT 647
 QY 101 RPI-----SHTPSVCDSELSFGSRSEGASGAPRWAKGDAAKRLALLWTPP 145
 Db 648 SAVTTGQHNITSSSTSMRLPSSNPETLSPFSDNSTSMPLLTSAHPTGGENITQVTP 707
 QY 146 -----PTPR-----GSHSPREPREA-----PLRAIHAPGSPS----- 170
 Db 708 ASISTHVSSTSPAPRGTTTSQASGPNSTSTKPGEVNVTGKTPONATSPQAPSGQKT 767
 QY 171 -----KTEPGPADSQKLSMGLHSSRLPKRGLSHLTHLVNPT---GHPATSAPTNGP 223
 Db 768 AVPTVTSTGGKANS---TTGGKHT-----GHGARTSTPTDYGGDSTTPRPNAT 817
 QY 224 QDLRPSGVS-----TFERSPLVTSRARSVSISVPSTPR 256
 Db 818 TYLEPSTSSKLPRWTFTSPVTTAQATVPVPTSPQPR 855
 RESULT 13
 CCAA_MOUSE
 ID_CCAA_MOUSE STANDARD; PRT; 2164 AA.
 AC P97435;
 DT 15-JUN-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
 DE channel, I type, alpha-1 polypeptide isoform 4) (Brain calcium
 DE channel I) (BI).
 GN CACNALA OR CACNL1A4 OR CCHA1A OR CACH4 OR CACN3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT TG LEU-601.
 RC STRAIN=DBA/2J;
 RX MEDLINE=97083572; PubMed=8929530;
 RA Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,
 RA Hawkes R., Frankel W.N., Copeland N.G., Jenkins N.A.;
 RT "Absence epilepsy in tottering mutant mice is associated with calcium
 RT channel defects."
 RL Cell 87:607-617(1996).
 CC -!- FUNCTION: Voltage-sensitive calcium channels (VSCC) mediate the
 CC entry of calcium ions into excitable cells and are also involved
 CC in a variety of calcium-dependent processes, including muscle
 CC contraction, hormone or neurotransmitter release, gene expression,
 CC cell motility, cell division and cell death. The isoform alpha-1A
 CC gives rise to P and/or Q-type calcium currents. P/Q-type calcium
 CC channels belong to the "high-voltage activated" (HVA) group and

are blocked by the funnel toxin (Ftx) and by the omega-agatoxin-
 IVA (omega-Aga-IVA). They are however insensitive to
 dihydropyridines (DHP), and omega-conotoxin-GVIA (omega-CTX-GVIA).
 -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE POSE-
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
 -!- SUBCELLULAR LOCATION: Integral membrane protein.
 -!- TISSUE SPECIFICITY: Brain specific; mainly found in the
 cerebellum, olfactory bulb, cerebral cortex, hippocampus, and
 inferior colliculus. In the hippocampus, expression occurs in
 pyramidal and granule neurons, as well as in interneurons.
 Purkinje cells contain predominantly P-type VSCC, the Q-type being
 a prominent calcium current in cerebellar granule cells.
 -!- DOMAIN: Each of the four internal repeats contains five
 hydrophobic transmembrane segments (S1, S2, S3, S5, S6) and one
 positively charged transmembrane segment (S4). S4 segments
 probably represent the voltage-sensor and are characterized by a
 series of positively charged amino acids at every third position.
 -!- DISEASE: DEFECTS IN CACNALA ARE THE CAUSE OF A DELAYED-ONSET,
 RECESSIVE NEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,
 RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE
 SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN
 HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)), THAT IS
 CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST WEANING,
 BUT NO MOTOR SEIZURES; AND ROLLING NAGAYA (TG(ROL)), THAT PRESENTS
 AN INTERMEDIARY PHENOTYPE, THE ATAXIA BEING SOMEWHAT MORE SEVERE
 THAN WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION
 OF CEREBEULAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF
 MUTANTS.
 -!- SIMILARITY: Belongs to the calcium channel alpha-1 subunits
 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U76716; AAC52940.1; --
 MGI; MGI:109482; Cnacnla.

GO; GO:0005891; C:voltage-gated calcium channel complex; IC.
 InterPro; IPR001682; Ca/Na pore.

InterPro; IPR002077; Ca channel alpha.

InterPro; IPR002111; Cat channel_TrgpL.

InterPro; IPR005821; Ion_trans_nlg.

InterPro; IPR005820; M-channel_nlg.

Pfam; PF00520; Ion_trans; 4.

PRINTS; PR0167; CCHANNEL.

PRINTS; PR01632; PVDCCALPHA1.

Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;

Calcium channel; Glycoprotein; Repeat; Multigene family;

Calcium-binding; Phosphorylation; Disease mutation.

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

REPEAT 427 671

REPEAT 1134 1417

REPEAT 1454 1717

REPEAT 17 317

DR GO; GO:0005622; C:intracellular; ISS.
DR GO; GO:0016192; P:vesicle-mediated transport; ISS.
FT NON TER 1 1
FT DOMAIN 67 75 POLY-GLU.
FT DOMAIN 459 462 POLY-GLU.
SQ SEQUENCE 513 AA; 55721.MW; 70CD36ACA49CB061 CRC64;
Query Match 8.5%; Score 122; DB 1; Length 513;
Best Local Similarity 23.8%; Pred. No. 0.45;
Matches 58; Conservative 22; Mismatches 98; Indels 66; Gaps 10;
QY 50 PDFPPW---VEKANRTRGVGKEASKALGKGC-----ETTPSRGSGTPTLTPR 95
DB 4 PRKPPMITLVOTEPKKPAQPPSSGFGPLSQAYRQVEDGLEEQTKSSGTEPKP- 62
QY 96 KKNYRPISTHPSYCDSELSFGSRSEGASFGAPRMAKGAALRALLMTPTPTPRGSHSPR 155
DB 63 ---YNP-----FEEEEEERGFAPPVFPSPSLAPPVFPSPAPFVP--SPAPA 104
QY 156 PREAPLRAIH-----PAGRSKTEPGPAADSQKLSMGLSHSRPLKGLSHSLTHLNVP 209
DB 105 PSEATPKSLHWYGITTSPPKTKRPAAPRAPSAPLAIHAR-----LSHSEPPS 155
QY 210 -TGHPA-----TSAPHTNGPDLPSTSGVTFRSPLVTSRARSVSI SVPTSPRRGGA 260
DB 156 ATPSPALSVESLSSESSHTANAPLEP-----PAVFKSSDDPAHVPGTPTGSGN 206
QY 261 TQKP 264
DB 207 SVTP 210
RESULT 15
YK82 MYCTU
ID _YK82 MYCTU STANDARD; PRT; 721 AA.
AC Q10690;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Rv2082.
GN Rv2082 OR MTCY49.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STEAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: SOME, TO M.TUBERCULOSIS Rv0029.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73966; CAA98194.1; -.
DR PIR; E70766; Rv2082.
DR TubercuList; Rv2082; -.

KW Hypothetical protein; Complete proteome.
FT DOMAIN 295 307 ALA-RICH.
SQ SEQUENCE 721 AA; 73564 MW; 28CAC21029ED0C57 CRC64;
Query Match 8.4%; Score 121; DB 1; Length 721;
Best Local Similarity 20.7%; Pred. No. 0.78;
Matches 78; Conservative 26; Mismatches 130; Indels 142; Gaps 14;
QY 2 KTVPELVASGMOTLQHRCRGY-----RVKARTSYVDTELFG----- 40
DB 122 KGPIEAKVAANVANIEQSNAMADHVGATAMGNIIDATQRFDETIGDHTWLRDHGVS 181
QY 41 -SPACTRTTPDPDFPPVWEKANRTRGVGKEASKALGKGCETTPSRGSGTPTLTPRKKNK 99
DB 182 DTFA--RRPVTAEDMTSMAN-----SPAGSPFGAAPSAPSHSTTTSGPTATTP 230
QY 100 YRPISHTPSYCDSELSFGSRSEG-----ASFQAPRMAKGAALRALLMTPTPTPRGS 151
DB 231 TSPFGTAEMVLSSS---STSSGPPPTAPTPTSPFGTAPMPG-----PPPGTV 275
QY 152 HSPRPREAPLRAIHGAPGSKTEPG-----PAADSO 181
DB 276 SPFLPPSAP--AVGVGSPVPAAGVPPAAAAATAPLSFQSLGQSFTHGTTGTPAAAG 333
QY 182 KLSMGLH--SSRPL-----KRLSHSLTHLNVP----- 209
DB 334 ALSAGALHAATEPLPPAPPPTPTVTVTATATTAGIPIIPDSAPTESPAPIAPTTD 393
QY 210 -----TCHPATSAHTNGP-----QDLRSTSGVTFRSPLVTSRARSVS 248
DB 394 NASAMTPIAPMVANGPPAPPAAPAAAPAGPLPAYGADLRPPVTTPPTPTGPISGAA 453
QY 249 ISVPSTPRRGATQKP 264
DB 454 VT-PSSPAAGSLMSP 468
Search completed: August 10, 2004, 12:11:35
Job time : 14 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1437	100.0	269	4	Q96K30	Q96K30 homo sapien
2	867.5	60.4	253	11	Q9D1H0	Q9D1H0 mus musculus
3	863	60.1	177	4	Q96K25	Q96K25 homo sapien
4	150	10.4	647	10	Q8S148	Q8S148 cryza sativ
5	144.5	10.1	518	5	Q8NQ38	Q8NQ38 caenorhabdi
6	144.5	10.1	524	5	Q02123	Q02123 caenorhabdi
7	144.5	10.1	539	5	Q8WQ39	Q8WQ39 caenorhabdi
8	142.5	9.9	964	5	Q28963	Q28963 trypanosoma
9	136	9.5	769	10	Q9LL41	Q9LL41 chlamydomon
10	134	9.3	1131	5	Q9VLW2	Q9VLW2 droscophila
11	133.5	9.3	1303	13	Q7SY42	Q7SY42 brachydanio
12	133	9.3	862	11	Q9JIK1	Q9JIK1 rattus norv
13	133	9.3	1240	12	Q3DMH8	Q3DMH8 rat cytomeg
14	132.5	9.2	488	16	Q87W07	Q87W07 pseudomonas
15	132.5	9.2	506	12	Q69125	Q69125 human herpes
16	132.5	9.2	580	4	Q9UF83	Q9UF83 homo sapien

```

DR EMBL; AK075358; BAC11568.1; -.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 28619 MW; 3F053BE1454F60773 CRC64;

Query Match
  100.0%; Score 1437; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 5.5e-107;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPELVAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFDPWVEKA 60
DB 1 MKTPELVAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFDPWVEKA 60

QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
DB 61 NRTGVGKEASKALGAKGSCETTPSRGSTPLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120

QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADS 180
DB 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADS 180

QY 181 QKLSMGGHLSRPLKRGHLSLTHLVNSTGHPATSPHTNGPQDLRPSSTGVTFRSPLV 240
DB 181 QKLSMGGHLSRPLKRGHLSLTHLVNSTGHPATSPHTNGPQDLRPSSTGVTFRSPLV 240

QY 241 TSARSVSISVPSTPRGGATOKPPPWK 269
DB 241 TSARSVSISVPSTPRGGATOKPPPWK 269

RESULT 2
Q9D1H0 PRELIMINARY; PRT; 253 AA.
AC Q9D1H0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 110008J03Rik protein (RIKEN CDNA 110008J03 gene).
GN 110008J03Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK003581; BAB22871.1; -.
DR EMBL; BC021365; AAB21365.1; -.

```

```

DR MGI:1922021; 1110008J03Rik.
SQ SEQUENCE 253 AA; 27070 MW; 239EC9BD9CF4FA61 CRC64;

Query Match
  60.4%; Score 867.5; DB 11; Length 253;
Best Local Similarity 66.3%; Pred. No. 1.5e-61;
Matches 171; Conservative 15; Mismatches 67; Indels 5; Gaps 1;

QY 12 MOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFDPWVEKANTRGVGEAS 71
DB 1 MKALHLQHRSPSYRVKARASYVDETLFGSPARTRPAQDFDPWVQVNCNRSRGVGP 60

QY 72 KALGAKGSCETTPSRGSTPLTPRKKNKYRPISTHTPSYCDLSFGSRSEGASFGAPRM 131
DB 61 KGLAKRDCESPSRSGSTPLTPRKKNKYRPISTHTPSYCDLSFGSKESG-----RM 115

QY 132 GDAAKLRALLWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGGAADSOKLSMGG 191
DB 116 GDAAKLRLLFWTPPTPRGSHSPRPREAPLRAIHPAGPSKTEPGTRPVRTAGS 175

QY 192 RPLKEGLSHSLTHLVNSTGHPATSPHTNGPQDLRPSSTGVTFRSPLVTSRAESV 251
DB 176 CSLGQRSHSLTHLVNSTGHPASSAQTNGPNSPRTNSGATVQSPLVTSKACSG 235

QY 252 PSTPRGGATOKPPPWK 269
DB 236 PAPPRRGACPPKPPPWK 253

RESULT 3
Q96K25 PRELIMINARY; PRT; 177 AA.
AC Q96K25;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14835.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027741; BAB55333.1; -.
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 19050 MW; 834FD0C7095AED0B CRC64;

Query Match
  60.1%; Score 863; DB 4; Length 177;
Best Local Similarity 99.4%; Pred. No. 2.3e-61;
Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTPELVAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFDPWVEKA 60
DB 1 MKTPELVAVSGMOTLGLQHRGCGYRVKARTSYVDETLFGSPAGTRTPDPDFDPWVEKA 60

QY 61 NRTGVGKEASKALGAKGSCETTPSRGSTPLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120
DB 61 NRTGVGKEASKALGAKGSCETTPSRGSTPLTPRKKNKYRPISTHTPSYCDLSFGSRSE 120

QY 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPL 161
DB 121 GASFGAPRMAGDAAKLRALLWTPPTPRGSHSPRPREAPV 161

```


SQ SEQUENCE 964 AA; 101611 MW; 8F0600289FB847AD CRC64;

Query Match 9.9%; Score 142.5; DB 5; Length 964;
 Best Local Similarity 26.8%; Pred. No. 0.0053;
 Matches 84; Conservative 24; Mismatches 109; Indels 97; Gaps 19;

QY 13 QTLGLQRCGGYR-----VKARTSYVDLTFGSPA-----43
 DB 609 RTPDISHYVGGYRSDMPISHTVNNVLLYNQLNAEIRTLFISQDLIGTEAHWDSS 668
 QY 44 -----GTRPTPPDPDPWVEKANRTRGV---GKEASKALGAGKGCETTPS-----RG 87
 DB 669 SDSNAHSTPSTPGD-----SSAHSSTPSTPADNGAHSTPSTPADNGAHSTPSTPGDNGAH 722
 QY 88 STPTLTTRKKNKYRPISTHSYCDSESLFGSRSEGAGPMAKGDAAKLALLWTPPT 147
 DB 723 STP-LTPADNGAH-----STPTPADSSAHSSTPSTPADNGA-----HSTPST 763
 QY 148 P-RGSHSPRPREAPLRAIHPA--GPSKTEPGPAADSQKLSM-----GGLHS--GRPLK 195
 DB 764 PADNGAHS-----TPLT---PADNGAHSTPTTP-ADSSAHSSTPSTPADNGAHSTPSTPAD 814
 QY 196 RGLSHSLTHLVNPTGHPATGAPNTNGPD--LRPSTSGVTFRPLVTSRARSVSISVPS 233
 DB 815 NG-AHSTPTPADSSAHSSTPSTPGDNGAHSTPSTPADNG-AHSTPL--TPADSSAHSSTPS 870
 QY 254 TPRGGATQKPKPP 267
 DB 871 TPGDNGAHSTPSAP 884

RESULT 9
 Q9LLJ1 PRELIMINARY; PRT; 769 AA.

AC Q9LLJ1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE CALK protein.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN MEDLINE=20379056; PubMed=10807915;
 RX Pan J., Snell W.J.;
 RA "Regulated targeting of a protein kinase into an intact flagellum. An
 RT aurea/iplp-like protein kinase translocates from the cell body into
 the flagella during gamete activation in Chlamydomonas."
 RL J. Biol. Chem. 275:24106-24114(2000).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF199021; AAF97501.1; -.
 DR HSSP; Q63450; 1A06.
 DR GO; GO:0005524; P:ATP binding; IEA.
 DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR002290; Ser thr kinase.
 DR InterPro; IPR008271; Ser_thr_kin_AS.
 DR Pfam; PF00069; pkinase; I.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 769 AA; 80250 MW; DB037FFCD4245101 CRC64;

Query Match 9.5%; Score 136; DB 10; Length 769;
 Best Local Similarity 22.8%; Pred. No. 0.013;
 Matches 75; Conservative 41; Mismatches 115; Indels 98; Gaps 15;

QY 32 SYVDLTFG-----SPAGTRTPPPDPPP-----WVEKANR-----62
 DB 323 SVEPRQLGFGSNAPGAGEPSGPKPEPRPLQVSSVAGGFLGHSSMSAMDCVRYRD 382
 QY 63 -----TRGVGKEASKAL-GAGKGCETTPSRGSTTTLTPRKKNK-----VRPISH 105
 DB 383 IHLDSVTGVMASAAATEALKSVKRVSDGDTTGGPRPSAAGGRVSRAGTARNSTGQAPLSR 442
 QY 106 TSYCDESLFGSRSE-----GASFGAPRMAKGDAAKLALLW-----TPPTPRGS--- 151
 DB 443 MNMSAQAVAAAAAAMASGSGFGGAHGEDSPKGRALTREELYASITAPLGGGSPVL 502
 QY 152 -----HSRPRE-AFLRAIHPAGPKTEPGPAADSQKLSMGLHSHSRPKRG---LS 199
 DB 503 ARSFTAGRASPAQQWAPL-----PTNACKRSGGGGLSGRSLSQATCATCPPSRGAELA 558
 QY 200 HSLTHLVNPTGSH-----PATSA-----PHTNG-----PQDLRPSISGVTFRSP 238
 DB 559 HDLSRLRLVDSAAASGGGGAAGPSTSGAVPYSPGKSSGIRPSPPOAPSPRTARGLTGSP 618
 QY 239 LVTSRARSVSISVPSTPRRGGATQKPKPP 267
 DB 619 LTPSL-SQSFTSPVSPPLAAGSNASAGTP 646

RESULT 10
 Q9VLW2 PRELIMINARY; PRT; 1131 AA.

AC Q9VLW2
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE CG14535 protein.
 DE CG14535.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galte R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brodeur G., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chery J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Malshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassaman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Gelniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banon J., An H., Baldwin D., Bonzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminck J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan J., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Gray A., Harris N.,
RA Krommiller B., Marshall B., Mallburn G., Richter J., Russo S.,
RA Searle S.M., Smith E., Sau S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003619; AAF52589.2;
DR FlyBase; FBgn0031955; CG14535.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IP0001752; P:kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1131 AA; 120721 MW; CB8413CD0F79466C CRC64;
Query Match
Best Local Similarity 27.58; Score 134; DB 5; Length 1131;
Matches 70; Conservative 30; Mismatches 99; Indels 56; Gaps 13;
QY 2 KTFVELAVSGNQTGLQHCRCG-GYRVKARTSYVDLFGSP---AGTRTPDFDPPMV 57
DB 402 RVFPLAVGLAQLGLGNGSSAGSDPSSEISADTVIYMGPNDDATDGEHPVPLSLT 461
QY 58 EKANRTGVCKEASKALGKGS-CETTPSR--GSTPTLTPR---KQKVRPISHTPSYC 110
DB 462 --AGDNRGV---MSKAL--KSGLEKFPKSAAGNSNMKMAAABAKKLPCHSGTSL- 513
QY 111 DESLFGSRSGASTGAPRMAKGAALALLWTTPPTPR-----GSHSPRPAPRAIATH 165
DB 514 -----KQAGAGACSSPMIPHEQPQIQ-A-MGSPFIPFRHVMVSGMVPSPKSPNRAH 566
QY 166 P-----AGPSKTEPGPADSQKLSMGLHSSRLKRLSH----- 200

DB 567 PGAALEQLEAGKVEKITEEOWIDGPRVSRKAVAEARHLMREVNHVQCETWVDGPKSQSQR 626
QY 201 SLTHLNVPTGHPAT 215
DB 627 SLTACNLPAAGGSQT 641
RESULT 11
Q7SY42 PRELIMINARY; PRT; 1303 AA.
AC Q7SY42;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein
OS *Brachydanio rerio* (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Heide F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055134; AAH55134.1; --
KW Hypothetical protein.
SQ SEQUENCE 1303 AA; 147812 MW; F10EF49F81EA7E17 CRC64;
Query Match
Best Local Similarity 9.3%; Score 133.5; DB 13; Length 1303;
Matches 58; Conservative 31; Mismatches 110; Indels 41; Gaps 8;
QY 50 PFDPPVWEKANRTGVCKEASKALGKGS-CETTPSRGSTPTLTPRKKN----- 98
DB 498 PSDKPAWAKVEERSKLNQSSPALQHK-----VANRISDFSLPFRSESGGMQPART 552
QY 99 --KYRPISTHSYCDSELFGRSESGASFGA-----PRWAKGDAKRALLTPTPTPRGS 151
DB 553 PMHRPVEPQVAH-----LVFVRSHSSNSVSVQSLHDSAGKVSFAQESLVSNERPQRN 609
QY 152 HSP---RPREAPLRAHPAGSKTEPGAADSQKLSMGLHSSRLKRGHSLHNLVNP 208
DB 609 SDPTSEMPAPRPMASHEEKPDSSPWLREDAVPPKLSGTTQVFPQRTTISPALARKNSP 668
QY 209 ST-----GHPTATAPHTNGPDILRPSTGVTFRSPLVTSRARSVSVSTPRRGATQ 262
DB 669 NGSSSGGGRASQLIRTSNP-DLRRTELSL-----DVALQRTSSNSSSSSSTSSQGGSQ 723

RESULT 12		J. Virol. 74:7656-7665 (2000).	
Q9JIK1	PRELIMINARY;	862 AA.	
AC Q9JIK1;			
DT 01-OCT-2000 (TRENBLrel. 15, Created)			
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT 01-OCT-2000 (TRENBLrel. 25, Last annotation update)			
DE Mu-protocadherin.			
OS Rattus norvegicus (Rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX NCBI_TaxID=10116;			
[1]			
SEQUENCE FROM N.A.			
RP STRAIN=Sprague-Dawley;			
EX MEDLINE=20387356; PubMed=10801787;			
RA Goldberg M., Peshkovsky C., Shifteh A., Al-Awqati Q.;			
RT "mu-protocadherin, a novel developmentally regulated protocadherin			
RT with mucin-like domains.";			
RL J. Biol. Chem. 275:24622-24629 (2000).			
DR EMBL; AF221952; AAF70456.1; -.			
GO GO:0016020; C:membrane; IEA.			
GO GO:0005509; F:calcium ion binding; IEA.			
GO GO:0007156; P:homophilic cell adhesion; IEA.			
DR InterPro; IPR002126; Cadherin.			
DR PRINTS; PRO0205; CADHERIN.			
DR SMART; SMO0112; CA; 3.			
DR PROSITE; PS00232; CADHERIN_1; 1.			
DR PROSITE; PS0268; CADHERIN_2; 3.			
SQ SEQUENCE 862 AA; 90976 MW; 302E302503A9C375 CRC64;			
Query Match		9.3%; Score 133; DB 11; Length 862;	
Best Local Similarity		27.7%; Pred.No. 0.027;	
Matches		66; Conservative 22; Mismatches 82; Indels 68; Gaps 14;	
QY 42	PAGTRPTPPDPDPWVEKANTRGVGKEASKALGAKGSCETTPSRGSGTPTLTTPRKKNYR	101	
DB 460	PSTESPTTPEAGGTTGSSNTT-----LETSTSTSGTSGPATSSGGAGFPFP-AGTTLS	514	
QY 102	PISHTPSYCDLSFGSRSEGFAGPRMAK--GDAKLRALLMTTP-----PTPRGS	151	
DB 515	PLTSAPTVPG---GSPTLGIS-TSPQTATPGGATQ-----TPKPGTSQPMVPTFGAS	563	
QY 152	HSPRPRAPLRAHPAGSKTEPGAPDAADSKLSMGLHSSRPLKXGLSHSLTHLNVPTG	211	
DB 564	TSSQFAT-----PSGSSTQTPKPG-----TSQPM-----VPTPG	592	
QY 212	HPATSAPHT---NGPQDLRPSGTGTFPSPLVTSRARSVSISVPSPTPRRGATQKPKP	266	
DB 593	ASTSSQATPSGSSTQTPRPTS-----QPMVPTFGASTS-SQATP--SGSTQTPKP	642	
RESULT 13			
Q9DWH8	PRELIMINARY;	1240 AA.	
AC Q9DWH8;			
DT 01-MAR-2001 (TRENBLrel. 16, Created)			
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)			
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)			
DE R2.			
GN R2.			
OS Rat cytomegalovirus (strain Maastricht).			
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC Betaherpesvirinae; Muromegalovirus.			
OX NCBI_TaxID=79700;			
[1]			
SEQUENCE FROM N.A.			
RP STRAIN=Maastricht;			
EX MEDLINE=20366325; PubMed=10906222;			
RA Vink C., Beuken E., Bruggeman C.A.;			
RT "Complete DNA sequence of the rat cytomegalovirus genome.";			
Query Match		9.2%; Score 132.5; DB 16; Length 488;	
Best Local Similarity		24.0%; Pred.No. 0.04;	
Matches		74; Conservative 19; Mismatches 115; Indels 100; Gaps 13;	
QY 42	PAGTRPTPPDPDPWVEKANTRGVGKEASKALGAKGSCETTPSRGSGTPTLTTPRKKNYR	101	
DB 250	PAAPPPPPQTPQPSVAPPHRRPSSSTLASLTFDGA-GVGVGLGPTFTAPGTPPVYQ	308	
QY 102	PISHTPSYCDLSFGSRSEGFAGPRMAKDA-----KLRALL	141	
DB 309	QPPSTAIY-----GAAAGVGFADGTAGGAEDDFGDEADLEGRRVFDOLAVV	357	
QY 142	WT-----PPTPRGSHSRPREAPL-----RAHPAG-----PSKTEPGPAA	178	
DB 358	RSDDLPPNHPGSLDVLPAAPTLDRPQVSPASRGDSLSRGAAGTWDPSVPEAPAG	417	
QY 179	DSQ-----KLSMGLHSSR-----PLKRG-----LSHS	201	
DB 418	PRAAGWPGTSSRG--SSRSCPGRLSDATPATPPASRGDASTEGTTTTTTTTTTTT	475	
QY 202	LTHLNVPTGCHPATSAHTNGPDQLRPSGTGTFPSPLVTSRARSVSISVPSPT-PRRGA	260	
DB 476	TTGLOGKGTGTVLT-PRTS-----APSTTATSAPTLPSTARSRCGSGTTFGSPRAR	530	
QY 261	TQPKPPW	268	
DB 531	TSTPDPW	538	
RESULT 14			
Q87W07	PRELIMINARY;	488 AA.	
ID Q87W07;			
AC Q87W07;			
DT 01-JUN-2003 (TRENBLrel. 24, Created)			
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)			
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE Type III effector Hoppmai (Pto).			
GN HOPPMai (Pto) OR PSPT04776.			
OS Pseudomonas syringae (pv. tomato).			
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC Pseudomonadaceae; Pseudomonas.			
OX NCBI_TaxID=323;			
[1]			
SEQUENCE FROM N.A.			
RP STRAIN=BC3000;			
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,			
RA Barry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,			
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,			
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,			
RA White O., Fraser C., Collier A.;			
RT "Complete sequence of Pseudomonas syringae";			
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AS016873; AAO58206.1; -.			
DR TIGR; PSPT04776; -.			
DR InterPro; IPR001623; DnaJ_N.			
DR Pfam; PF00226; DnaJ; 1.			
DR PROSITE; PS0076; DnaJ_2; 1.			
KW Complete proteome.			
SQ SEQUENCE 488 AA; 52549 MW; 4D402A03A3696DDF CRC64;			

Qy	851	GGCGTGGGCAAGGAGGCATCGAAGGCCTTGGGGCAAGGGGAGCTGTGAGACCAACCC	910
Db	661	GGCGTGGGCAAGGAGGCATCGAAGGCCTTGGGGCAAGGGGAGCTGTGAGACCAACCC	720
Qy	911	TCRAGGGGCGACCCACCCCTCACACCAAGGAGGAGACAAATACAGACCCATCAGC	970
Db	721	TCRAGGGGCGACCCACCCCTCACACCAAGGAGGAGACAAATACAGACCCATCAGC	780
Qy	971	CACACCCCGTCTTACTGTGATGAGTCGCTGTGTTGGCTCCGATCTGAAGCGCCAGCTTC	1030
Db	781	CACACCCCGTCTTACTGTGATGAGTCGCTGTGTTGGCTCCGATCTGAAGCGCCAGCTTC	840
Qy	1031	GGGGCCCGGGGATGGCGAAGGGGGATGCCGCAAGCTCCGTGCTCTCTTGTGGACGCCA	1090
Db	841	GGGGCCCGGGGATGGCGAAGGGGGATGCCGCAAGCTCCGTGCTCTCTTGTGGACGCCA	900
Qy	1091	CCACTACCCCCAGGGGTAGCCACTCGCCCCCGCCACAGGAGGACCACTGCCAGGCCATT	1150
Db	901	CCACTACCCCCAGGGGTAGCCACTCGCCCCCGCCACAGGAGGACCACTGCCAGGCCATT	960
Qy	1151	CACCGAGCTGGTCCCTCAAGAAGAGCCGGGGCCAGGGGAGACTCCCAAGAGTTATCT	1210
Db	961	CACCGAGCTGGTCCCTCAAGAAGAGCCGGGGCCAGGGGAGACTCCCAAGAGTTATCT	1020
Qy	1211	ATGGGTGGGTTACACTCTTCACGGCCCTGAAGGGGGACTTTCCCAATCCCTCACCCAC	1270
Db	1021	ATGGGTGGGTTACACTCTTCACGGCCCTGAAGGGGGACTTTCCCAATCCCTCACCCAC	1080
Qy	1271	CTGAATGTCCCCAGCACTGTGTCATPCAGACCAAGTGCCTCCCAACAAATGGGCTCTCAG	1330
Db	1081	CTGAATGTCCCCAGCACTGTGTCATPCAGACCAAGTGCCTCCCAACAAATGGGCTCTCAG	1140
Qy	1331	GATCTCAGGCGCTTCCAGTCAGGGGTGACCTTCGGAGCCGCCCTGTGTGACTTCCAGGGCT	1390
Db	1141	GATCTCAGGCGCTTCCAGTCAGGGGTGACCTTCGGAGCCGCCCTGTGTGACTTCCAGGGCT	1200
Qy	1391	CGCTCAGTTAGCATTTTCAGTGCCATCTACCCCCACGACGAGGTGGGGCCACCCAGAAACCA	1450
Db	1201	CGCTCAGTTAGCATTTTCAGTGCCATCTACCCCCACGACGAGGTGGGGCCACCCAGAAACCA	1260
Qy	1451	AAGCCCGCTTGGAAATGATACTCTTTTCATCAGGGTTGCCTATGGGGCCACGGCGACAGGT	1510
Db	1261	AAGCCCGCTTGGAAATGATACTCTTTTCATCAGGGTTGCCTATGGGGCCACGGCGACAGGT	1320
Qy	1511	ATGGCCCGCTTGCAGGGTAGGAGACATTCATACCCAGGGAAACCCAGAGTATTAAAGAA	1570
Db	1321	ATGGCCCGCTTGCAGGGTAGGAGACATTCATACCCAGGGAAACCCAGAGTATTAAAGAA	1380
Qy	1571	GCCTCTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCTTCCTTTATCTGTACAA	1630
Db	1381	GCCTCTGGGGGAGACAGACATAGCAGGGGTGGGAGTGCTTCCTTTATCTGTACAA	1440
Qy	1631	TCTCTAGTCGATCTTCCCTTTTCTCCGATTCGCGATTTGGGGGCCACTCTAAGATG	1690
Db	1441	TCTCTAGTCGATCTTCCCTTTTCTCCGATTCGCGATTTGGGGGCCACTCTAAGATG	1500
Qy	1691	CTCTCTCCAGCCCTGTCTCAACATATCTCAAAATTAGTGCACCCAGGGGCTGTGCAC	1750
Db	1501	CTCTCTCCAGCCCTGTCTCAACATATCTCAAAATTAGTGCACCCAGGGGCTGTGCAC	1560
Qy	1751	CTCCACATCATCATGTCCTGTGTCGCAAGTGCGAATAAACGGCGTGATTGCCAACCTG	1810
Db	1561	CTCCACATCATCATGTCCTGTGTCGCAAGTGCGAATAAACGGCGTGATTGCCAACCTG	1620
Qy	1811	G 1811	
Db	1621	G 1621	
RESULT 6			
AK027733			
LOCUS			
DEFINITION Homo sapiens cDNA FLJ14827 fis, clone OVARC1000896.			
AK027733 1775 bp mRNA linear PRI 01-AUG-2002			

Db	781	CACACCCCGTCTTACTGTGTAGTAGTGCCTGTTTGGCTCCCGATCTGAAGGCGCCACGCTTC	840
Qy	1031	GGGGCCCCCGCGGATGGCGAAGGGGGGATGCCGCAAAAGTCGCTGCTCTCTTTGTGGAGCGCCA	1090
Db	841	GGGGCCCCCGCGGATGGCGAAGGGGGGATGCCGCAAAAGTCGCTGCTCTCTTTGTGGAGCGCCA	900
Qy	1091	CCACCTACCCCCCAGGGGTAGCCACTCGCCCGCGCCACAGGAGGAGCAACAATCGAGGCATTT	1150
Db	901	CCACCTACCCCCCAGGGGTAGCCACTCGCCCGCGCCACAGGAGGAGCAACAATCGAGGCATTT	960
Qy	1151	CACCCAGCTGGTCCCTCCAAGACAGAGCCGGGGCCAGCGGCAGACTCCAGAAAGTTATCT	1210
Db	961	CACCCAGCTGGTCCCTCCAAGACAGAGCCGGGGCCAGCGGCAGACTCCAGAAAGTTATCT	1020
Qy	1211	ATGGGTGGGTACACTCTTCACGCCCCCTGAAGGGGGGACTTTCCATTTCCCTCAACCCAC	1270
Db	1021	ATGGGTGGGTACACTCTTCACGCCCCCTGAAGGGGGGACTTTCCATTTCCCTCAACCCAC	1080
Qy	1271	CTGAATCTCCCCCAGCACTGGTCAATCAGCCACAGATGCCCCCACCACAAATGGGCTCAG	1330
Db	1081	CTGAATCTCCCCCAGCACTGGTCAATCAGCCACAGATGCCCCCACCACAAATGGGCTCAG	1140
Qy	1331	GATCTCAGGCGCTTCCACGTCAGGGGTGACCTTCGGAGCGCCCTGTFGACTTCCAGGGCT	1390
Db	1141	GATCTCAGGCGCTTCCACGTCAGGGGTGACCTTCGGAGCGCCCTGTFGACTTCCAGGGCT	1200
Qy	1391	CGTCAAGTTAGCAATTCAGTGCCATCTACCCACGACGAGGTGGGGCCACCCAGAAACCA	1450
Db	1201	CGTCAAGTTAGCAATTCAGTGCCATCTACCCACGACGAGGTGGGGCCACCCAGAAACCA	1260
Qy	1451	AAGCCCCCTTTGGAAATGATACTCTTTTCATCAGGGTTGCCCTATGGGGCCACGGCGACAGGT	1510
Db	1261	AAGCCCCCTTTGGAAATGATACTCTTTTCATCAGGGTTGCCCTATGGGGCCACGGCGACAGGT	1320
Qy	1511	ATGGCCCCCTTTGCCAGGGTAGGAGACATTCATACCCAGGGAACCCAGGTATTAAGAA	1570
Db	1321	ATGGCCCCCTTTGCCAGGGTAGGAGACATTCATACCCAGGGAACCCAGGTATTAAGAA	1380
Qy	1571	GCCCTCTGGGGGCAGACACATAGCAGGGGTGGGCAGTGCCTTCCCTTATCCTGCACAA	1630
Db	1381	GCCCTCTGGGGGCAGACACATAGCAGGGGTGGGCAGTGCCTTCCCTTATCCTGCACAA	1440
Qy	1631	TCTCTAGTCGATTTCTGCCCTTTTCTCCCGATTCGGGATTTGGGGGCCACTCTAAGATG	1690
Db	1441	TCTCTAGTCGATTTCTGCCCTTTTCTCCCGATTCGGGATTTGGGGGCCACTCTAAGATG	1500
Qy	1691	CCTCTCTCCAGCCCTGTCTCAACCATACTCCAATTAGTCCCAACCCAGGGGCTGGCAC	1750
Db	1501	CCTCTCTCCAGCCCTGTCTCAACCATACTCCAATTAGTCCCAACCCAGGGGCTGGCAC	1560
Qy	1751	CTCCACATCATCATCTCTTGTGTCGCAAGTCGAATAAACCGGCTGATTGCCAACCTTG	1810
Db	1561	CTCCACATCATCATCTCTTGTGTCGCAAGTCGAATAAACCGGCTGATTGCCAACCTTG	1620
Qy	1811	G 1811	
Db	1621	G 1621	
RESULT	5		
BD158346			
LOCUS	BD158346	1775 bp	DNA linear PAT 17-JAN-2003
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD158346		
VERSION	BD158346.1	GI:27864104	
KEYWORDS	JP 2002191363-A/13189.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 1775)		
	Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,		
	Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.		


```
QY 1079 TTGTGGACGACACCTACCCAGGAGGTAGCCACTCGCCCGCCCGCCAGGAGGACCA 1138
Db 901 TTGTGGACGACACCTACCCAGGAGGTAGCCACTCGCCCGCCCGCCAGGAGGACCA 960
QY 1139 CTGGAGGACATTCACCCAGCTGGTCCCTCCAGAGCAGAGCCGGGCGCAGAGCTCC 1198
Db 961 CTGGAGGACATTCACCCAGCTGGTCCCTCCAGAGCAGAGCCGGGCGCAGAGCTCC 1020
QY 1199 CAGAGGATTCATAGGGTGGGTACACTCTTCAGCCCGCCCTGAAGCCGGGACCTTCCCAT 1258
Db 1021 CAGAGGATTCATAGGGTGGGTACACTCTTCAGCCCGCCCTGAAGCCGGGACCTTCCCAT 1080
QY 1259 TCCCTCACCCACCTGAATGTCCCGAGCACTGGTCACTCAGCCACAGTGGCCCCCACCACA 1318
Db 1081 TCCCTCACCCACCTGAATGTCCCGAGCACTGGTCACTCAGCCACAGTGGCCCCCACCACA 1140
QY 1319 AATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACCTTCCGAGCCCGCTGGTG 1378
Db 1141 AATGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACCTTCCGAGCCCGCTGGTG 1200
QY 1379 ACTTCCAGGCTCGCTCAGTTAGCATTTCACTGCGCATCTACCCACAGCAGGTGGGGCC 1438
Db 1201 ACTTCCAGGCTCGCTCAGTTAGCATTTCACTGCGCATCTACCCACAGCAGGTGGGGCC 1260
QY 1439 ACCCAGAAACCAAAGCCCTTGGAAATGATACCTTTTCATCAGGGTGGCTATGGGGCC 1498
Db 1261 ACCCAGAAACCAAAGCCCTTGGAAATGATACCTTTTCATCAGGGTGGCTATGGGGCC 1320
QY 1499 ACGGCGACAGTATGGCCCTTGGCCAGGTAGGAGACATTCATCACCGAGGAAACCCCA 1558
Db 1321 ACGGCGACAGTATGGCCCTTGGCCAGGTAGGAGACATTCATCACCGAGGAAACCCCA 1380
QY 1559 GGTATTAAGAGAGCCCTGTGGGGGCGACAGACATAGCAGGGGTGGGCGAGTGCCTCCCT 1618
Db 1381 GGTATTAAGAGAGCCCTGTGGGGGCGACAGACATAGCAGGGGTGGGCGAGTGCCTCCCT 1440
QY 1619 TTATCTGACAACTCTAGTCGATTTTCGCTTTTCTCCGATTCGGATTTGGGGGCC 1678
Db 1441 TTATCTGACAACTCTAGTCGATTTTCGCTTTTCTCCGATTCGGATTTGGGGGCC 1500
QY 1679 ACCTTAAGATGCTCTCTCAGCCCTGTCTCAACCATATCTCCAAATAGTGGCAACCCA 1738
Db 1501 ACCTTAAGATGCTCTCTCAGCCCTGTCTCAACCATATCTCCAAATAGTGGCAACCCA 1560
QY 1739 GGGGCTGGACCTCCACATATCCATTTGCTTGGCTGCGAAGTGGCAATTAACGGGTG 1798
Db 1561 GGGGCTGGACCTCCACATATCCATTTGCTTGGCTGCGAAGTGGCAATTAACGGGTG 1620
QY 1799 ATTGCCACCTGG 1811
Db 1621 ATTGCCACCTGG 1633
```

RESULT 4

```
LOCUS AX880378 1775 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 15283 from Patent EP1074617.
ACCESSION AX880378
VERSION AX880378.1 GI:40035114
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K., and Otsuki, T.
TITLE Primers for synthesizing full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 15283 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..1775
```

CDS

```
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
459..1278
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAE90924.1"
/db_xref="GI:40035115"
/translation="MKTPELVAVSGMOTLGLQHRCGGVYRKARTSYVDETLFGSPAG
TEPTDPDPMVEKANEKTRGVGEAKALGKAGSCETTPSRGSPPTLTPRKKNYRP
ISHTSYCDESLFGSRSEGASFGAPRAKAGDAKLRALLWTPTPPRSHSRPREAP
LRAIHPAGSPKTEPGFAADSKLSMGUHSRPLKRGLSHSLTHLVNFTGHPATSAF
HTNGPDLPSTSGVTPRSPLVTSRVSISVPSPTPRRGATQKPKPPWK"
ORIGIN
Query Match 89.5%; Score 1621; DB 6; Length 1775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 GTGCAAGTGTGGGTCTCTGGGTCTTCTGGATTCGCGGCGCTTACAGCTAGCTGTGCC 250
Db 1 GTGCAAGTGTGGGTCTCTGGGTCTTCTGGATTCGCGGCGCTTACAGCTAGCTGTGCC 60
QY 251 GGCTCTCTGGGTGAGTTCGGTCCGCGCGGTGCCCGGAGCGGCTAGGCTGCGGGGGT 310
Db 61 GGCTCTCTGGGTGAGTTCGGTCCGCGCGGTGCCCGGAGCGGCTAGGCTGCGGGGGT 120
QY 311 CGGGGCGCCCGAGGCATTCGGGCTCGAGATTGACGGGATCCCGGATGACCGCGCGGCC 370
Db 121 CGGGGCGCCCGAGGCATTCGGGCTCGAGATTGACGGGATCCCGGATGACCGCGCGGCC 180
QY 371 CGCGGCTTCACCGAGCGGTCCAGCTGTGGTGGGAAGAGTGGGGGACGGGTCCCTGA 430
Db 181 CGCGGCTTCACCGAGCGGTCCAGCTGTGGTGGGAAGAGTGGGGGACGGGTCCCTGA 240
QY 431 GGATCCGATGCTACGAGCCAAAGTGTCTAGTGTGTAGTGTGACCTACACATGTGAC 490
Db 241 GGATCCGATGCTACGAGCCAAAGTGTCTAGTGTGTAGTGTGACCTACACATGTGAC 300
QY 491 TTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCAAGTGTCT 550
Db 301 TTCACTCAGTTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCAAGTGTCT 360
QY 551 GAGAGATTAATGAAGCAATGCTTTGAAAGCTCTTTCAGAGGAGGAGCTCGGAGCAG 610
Db 361 GAGAGATTAATGAAGCAATGCTTTGAAAGCTCTTTCAGAGGAGGAGCTCGGAGCAG 420
QY 611 GGCTTGGCGGCGAGAGCACACCTGCTGTACAGGAGCCACAGGAGCATGAAGACCCCC 670
Db 421 GGCTTGGCGGCGAGAGCACACCTGCTGTACAGGAGCCACAGGAGCATGAAGACCCCC 480
QY 671 GTGAGTGTGGCGTCAAGTGGGATGACACCTTGGGCTTCAAGCCGTGCGCGAGGTGGC 730
Db 481 GTGAGTGTGGCGTCAAGTGGGATGACACCTTGGGCTTCAAGCCGTGCGCGAGGTGGC 540
QY 731 TACCGGGTCAAGCGCAGGAGCATCGATATGATGATGATGATGATGATGATGATGATGAT 790
Db 541 TACCGGGTCAAGCGCAGGAGCATCGATATGATGATGATGATGATGATGATGATGATGAT 600
QY 791 ACCCGGCTTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAAAGGTAAACAGAACCA 850
Db 601 ACCCGGCTTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAAAGGTAAACAGAACCA 660
QY 851 GGGGTGGGCGAGGAGCATCGAAGCTTGGGGCAAGAGGAGCTGTGAGACCAACCCCC 910
Db 661 GGGGTGGGCGAGGAGCATCGAAGCTTGGGGCAAGAGGAGCTGTGAGACCAACCCCC 720
QY 911 TCAAGGGCGAGCACCCCGCCCTCACACCAAGGAGAGAAACAATACAGACCCATCAGC 970
Db 721 TCAAGGGCGAGCACCCCGCCCTCACACCAAGGAGAGAAACAATACAGACCCATCAGC 780
QY 971 CACACCCGCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1030
Db 1030 CACACCCGCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
```

```
QY 1321 TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGTGACCTCCGGAGCCCTCGTGAC 1380
Db 1321 TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGTGACCTCCGGAGCCCTCGTGAC 1380
QY 1381 TTCCAGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGAGGTTGGGCCAC 1440
Db 1381 TTCCAGGCTCGCTCAGTTAGCATTTTCAGTGCCATCTACCCACAGAGGTTGGGCCAC 1440
QY 1441 CCAGAACCAAGCCCTTGGAAATGATACCTTTTCATCAGGGTTGCCATGGGCCAC 1500
Db 1441 CCAGAACCAAGCCCTTGGAAATGATACCTTTTCATCAGGGTTGCCATGGGCCAC 1500
QY 1501 GGGCAGAGTATGGCCCTTCCAGGATAGGAGACATTCATCACCAGGAGAACCCAGG 1560
Db 1501 GGGCAGAGTATGGCCCTTCCAGGATAGGAGACATTCATCACCAGGAGAACCCAGG 1560
QY 1561 TATTAAAGAGCCCTTGTGGGGGAGACAGACATAGAGGGTGGGCAGTGCTTCCCTTT 1620
Db 1561 TATTAAAGAGCCCTTGTGGGGGAGACAGACATAGAGGGTGGGCAGTGCTTCCCTTT 1620
QY 1621 ATCTGACAACTCTAGTCGATTTTCCCTTTTCTCCCGATGCGGATTTGGGGGCCAC 1680
Db 1621 ATCTGACAACTCTAGTCGATTTTCCCTTTTCTCCCGATGCGGATTTGGGGGCCAC 1680
QY 1681 CTCTAAGATGCTTCTCCAGCCCTGTCTCAACCATACTCCAAATAGTGCCAAACCCAG 1740
Db 1681 CTCTAAGATGCTTCTCCAGCCCTGTCTCAACCATACTCCAAATAGTGCCAAACCCAG 1740
QY 1741 GGCTGGACCTCCCATCATCTGCTCTCTCCAGTGGCGATTTGGGGGCCAT 1800
Db 1741 GGCTGGACCTCCCATCATCTGCTCTCTCCAGTGGCGATTTGGGGGCCAT 1800
QY 1801 TGCCAACTCG 1811
Db 1801 TGCCAACTCG 1811

RESULT 3
AY007158 1839 bp mRNA linear PRI 31-AUG-2000
LOCUS Homo sapiens clone CDABP0113 mRNA sequence.
DEFINITION
ACCESSION AY007158
VERSION AY007158.1 GI:9956071
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Anderson, B., Wentland, M.A., Ricafrente, J.Y., Liu, W. and Gibbs, R.A.
A 'double adaptor' method for improved shotgun library construction
Anal. Biochem. 236 (1), 107-113 (1996)
96207227
PUBMED 8619474
REFERENCE 2 (bases 1 to 1839)
AUTHORS Yu, W., Anderson, B., Worley, K.C., Muzny, D.M., Ding, Y., Liu, W.,
Ricafrente, J.Y., Wentland, M.A., Lennon, G. and Gibbs, R.A.
Large-scale concatenation cDNA sequencing
Genome Res. 7 (4), 353-358 (1997)
97264341
PUBMED 9110174
REFERENCE 3 (bases 1 to 1839)
AUTHORS Zhou, J., Yu, W., Tang, H., Mei, G., Tsang, Y.T.M., Bouck, J., Gibbs, R.A.
and Margolin, J.F.
Direct Submission
Submitted (24-JUL-2000) Human Genome Sequencing Center and Texas
Children's Cancer Center, Baylor College of Medicine, Houston, TX
77030, USA
The clone request should be directed to Dr. J. Margolin at
Pediatrics-Hematology & Oncology, Texas Children's PEIGIN Center
102514, Houston, Texas 77030, USA. Telephone: 713-770-4583
margolinbcm.tmc.edu.
Location/Qualifiers
```

```
source 1. .1839
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CDABP0113"
/dev stage="infant"
/note="from patient with acute lymphoblastic leukemia"

ORIGIN
Query Match 90.1%; Score 1631.4; DB 9; Length 1839;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 CCTGTTCTCGGTGCAAAAGTCGTGGTTCCTGGGTTCTTGATTCGGGGCCGTTACAC 238
Db 1 CCTGTTCTCGGTGCAAAAGTCGTGGTTCCTGGGTTCTTGATTCGGGGCCGTTACAC 60
QY 239 GTAGCTGTGCGGCTCTCTCGGTGAGTCTCGTCCGCGGGTCCCGGGACCGGCTAG 298
Db 61 GTAGCTGTGCGGCTCTCTCGGTGAGTCTCGTCCGCGGGTCCCGGGACCGGCTAG 120
QY 299 GCTGCGGGGTTCGGGGCCCGAGGCATTCGGGCTGCAGATTCACGGGATCCGGATG 358
Db 121 GCTGCGGGGTTCGGGGCCCGAGGCATTCGGGCTGCAGATTCACGGGATCCGGATG 180
QY 359 CACGCGCGCCCCCGCGCCCTTACCGAGGGTCCAGACCTGTGTGGGAGAGGTCCGGGG 418
Db 181 CACGCGCGCCCCCGCGCCCTTACCGAGGGTCCAGACCTGTGTGGGAGAGGTCCGGGG 240
QY 419 ACGGTCCTCAGAGATCCGATGCTAGCAGCAGAGTGTCTCAGCTTTATAGGTGTGACC 478
Db 241 ACGGTCCTCAGAGATCCGATGCTAGCAGCAGAGTGTCTCAGCTTTATAGGTGTGACC 300
QY 479 TACACATGTGACTTCACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGCTACTT 538
Db 301 TACACATGTGACTTCACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGCTACTT 360
QY 539 CACAGTGTGTGAGAGGATTAATGAACAAATGCTTTGTAAGTCTTTTGAGAGGGAG 598
Db 361 CACAGTGTGTGAGAGGATTAATGAACAAATGCTTTGTAAGTCTTTTGAGAGGGAG 420
QY 599 COTCGGAAGCAGGCGCTGGCGGAGACACCTGTGTCCAGGAGACCCACAGGAGC 658
Db 421 COTCGGAAGCAGGCGCTGGCGGAGACACCTGTGTCCAGGAGACCCACAGGAGC 480
QY 659 ATGAAGACCCCGCTGGAGCTGGCGCTCAGTGGGATGCAGACCTCTGGGCTTTGACACCGC 718
Db 481 ATGAAGACCCCGCTGGAGCTGGCGCTCAGTGGGATGCAGACCTCTGGGCTTTGACACCGC 540
QY 719 TGCCGAGTGGTACCGGCTCAAGCCAGGACGTCATATGTGATGAGACTCTGTTGGC 778
Db 541 TGCCGAGTGGTACCGGCTCAAGCCAGGACGTCATATGTGATGAGACTCTGTTGGC 600
QY 779 AGCCGACAGGCACCCCGCCTTACCCACCGGACTTCGATCCGCCCTTGGTGGAAGGCT 838
Db 601 AGCCGACAGGCACCCCGCCTTACCCACCGGACTTCGATCCGCCCTTGGTGGAAGGCT 660
QY 839 AACAGAACAGAGGCGTGGGAGAGGAGCATCGAAGGCTTGGGGGCAAGAGGAGCTGT 898
Db 661 AACAGAACAGAGGCGTGGGAGAGGAGCATCGAAGGCTTGGGGGCAAGAGGAGCTGT 720
QY 899 GAGACACCCCTCTAAGGGGAGCAGCACCCACCCCTTACACCAAGGAAGAAACAATATAC 958
Db 721 GAGACACCCCTCTAAGGGGAGCAGCACCCACCCCTTACACCAAGGAAGAAACAATATAC 780
QY 959 AGACCCATAGCCACACCCCGTCTTCTGATGATGCTGTTTGGCTCCGATCTGAA 1018
Db 781 AGACCCATAGCCACACCCCGTCTTCTGATGATGCTGTTTGGCTCCGATCTGAA 840
QY 1019 GCGCGACCTTCGGGGCCCGGGATGCGGAGGGGATGCGCAAAAGCTCCGTGCTCTC 1078
Db 841 GCGCGACCTTCGGGGCCCGGGATGCGGAGGGGATGCGCAAAAGCTCCGTGCTCTC 900
```

```

RESULT 2
AK075358
LOCUS       AK075358               1811 bp      mRNA      linear      PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA P8EC0043 fis, clone NT2RP1001002.
ACCESSION   AK075358
VERSION     AK075358.1 GI:22761393
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1
AUTHORS     Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayaishi, K.,
            Ishii, S., Sato, K., Yamamoto, J., Wakamatsu, A., Negai, T.,
            Nakamura, Y., Nagahari, K., Sugano, S. and Isogai, T.
TITLE       HRI human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2
AUTHORS     Isogai, T. and Yamamoto, J.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT     HRI human cDNA sequencing project; cDNA 5' - 3' - end one pass
            sequencing, clone selection and full insert sequencing; Helix
            Research Institute (supported by Japan Key Technology Center etc.);
            cDNA library construction; Institute of Medical Science, University
            of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES             Location/Qualifiers
     source           1..1811
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP1001002"
                     /cell_lines="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP1"
     note="cloning vector: pUC19FL3-mRNA from NT2 neuronal
     precursor cells after 48-hours retinoic acid (RA)
     induction"
     CDS              659..1468
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC11568.1"
                     /db_xref="GI:22761394"
                     /translation="MKTPVELAVSGMOTLQHPRCRGYRVKARTSYVDETLFGSPAG
                     TRTPPPDPFWPEKANTRGVKEASKALGAKGSCSTPSRGSTPLTPRKKNKXP
                     ISHTPSYCDSELSFGSRGASFGAPRVAKGDAKRLALLWTPPTPRGSHSPRPAP
                     LRAIHPAGFSTKTEGPAADQKSMGLHSLRPLKRLGSLHLNVPSTGHPTASAP
                     HTNGPQDLRPSTSGVTRSPVLTSTRARSVSISVESTPRRGATQKPKPPWK"

ORIGIN
Query Match      100.0%; Score 1811; DB 9; Length 1811;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTGCTCGGCGCCCTCGGCGCCCTAGTACACACGACCTGAGTGAGGACAGGAC 60
DB 1 AGTGCTCGGCGCCCTCGGCGCCCTAGTACACACGACCTGAGTGAGGACAGGAC 60

QY 61 CCTTCCATGTTAGGACCTCTGGGCTCAGAGAGGTGGCGCCCGCCCTGGGCGGAC 120
DB 61 CCTTCCATGTTAGGACCTCTGGGCTCAGAGAGGTGGCGCCCGCCCTGGGCGGAC 120

QY 121 TCCCGCCATCCGCGGCGCGCAATGGTCGGGTCCGCTCCGAGTCTGCTGGTCTCCC 180
DB 121 TCCCGCCATCCGCGGCGCGCAATGGTCGGGTCCGCTCCGAGTCTGCTGGTCTCCC 180

QY 181 TGGTGTGGGTGCAAGTGTCTGGGTTCTGGATTTCGGGCGCGGCTTCACACGT 240
DB 181 TGGTGTGGGTGCAAGTGTCTGGGTTCTGGATTTCGGGCGCGGCTTCACACGT 240

```

```

QY 241 AGCTGTGCGGCTCTCTCGGCTGAGTCCGTCGCGCGCGGTGCTCCCGGAGCGGCTAGGC 300
DB 241 AGCTGTGCGGCTCTCTCGGCTGAGTCCGTCGCGCGCGGTGCTCCCGGAGCGGCTAGGC 300

QY 301 TGGCGGGGTCCGGGGCCCGCAGGCATTCGCGGCTGCAGATTGACGGGGATCCCGGATGCA 360
DB 301 TGGCGGGGTCCGGGGCCCGCAGGCATTCGCGGCTGCAGATTGACGGGGATCCCGGATGCA 360

QY 361 CGCGCGCCCCCGCGCCCTTACCGAGCGGTTCAGACCTGGTGGGAGAGAGGTTCGGGGAC 420
DB 361 CGCGCGCCCCCGCGCCCTTACCGAGCGGTTCAGACCTGGTGGGAGAGAGGTTCGGGGAC 420

QY 421 GGTGCTCTCAGGATCCCGATGCTCAGACGCAAGATGCTCAGCTTTATAGGTGTGACCTA 480
DB 421 GGTGCTCTCAGGATCCCGATGCTCAGACGCAAGATGCTCAGCTTTATAGGTGTGACCTA 480

QY 481 CACATGTGACTTCACTCAGTCTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCA 540
DB 481 CACATGTGACTTCACTCAGTCTTGTGATCCGTAATAATGGACAAATTCGAAGCTACTTCA 540

QY 541 CAGTGTGTTGAGAGATTAAATGAAACAATGCTTTGTAAGCTCTTTCGACGAGGAGCC 600
DB 541 CAGTGTGTTGAGAGATTAAATGAAACAATGCTTTGTAAGCTCTTTCGACGAGGAGCC 600

QY 601 TCGGAAGCAGAGGCTTGGCGCGCAGAGCACACCTGCTGTCCACGAGGACACACAGGAGCAT 660
DB 601 TCGGAAGCAGAGGCTTGGCGCGCAGAGCACACCTGCTGTCCACGAGGACACACAGGAGCAT 660

QY 661 GAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGAGACCTTCGCGCTTCAGCACCGGTG 720
DB 661 GAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGAGACCTTCGCGCTTCAGCACCGGTG 720

QY 721 CCGAGTGTCTACCGGCTCAAGGCCAGGACGTCATATGTGATGAGATCTGTGTTGGCAG 780
DB 721 CCGAGTGTCTACCGGCTCAAGGCCAGGACGTCATATGTGATGAGATCTGTGTTGGCAG 780

QY 781 CCAGCAGGACCCCGGCTTACCCCAACCGACTTCGATCCGCCCTGGTGGAGAAAGGCTAA 840
DB 781 CCAGCAGGACCCCGGCTTACCCCAACCGACTTCGATCCGCCCTGGTGGAGAAAGGCTAA 840

QY 841 CAGNACCAAGAGCGTGGCGCAGGAGGATCGAGGCTTGGGGGCAAGAGGAGGCTGTA 900
DB 841 CAGNACCAAGAGCGTGGCGCAGGAGGATCGAGGCTTGGGGGCAAGAGGAGGCTGTA 900

QY 901 GACCAACCCCTCAAGGGGCGAGCACCCCAACCTCACACCAAGAAAGAAAGAAATACAG 960
DB 901 GACCAACCCCTCAAGGGGCGAGCACCCCAACCTCACACCAAGAAAGAAAGAAATACAG 960

QY 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAGG 1020
DB 961 ACCCATCAGCACACCCCGTCTTACTGTGATGAGTGGCTGTTGGCTCCGATCTGAAGG 1020

QY 1021 CGCGAGTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCAAAAGCTCCGTGCTCTTT 1080
DB 1021 CGCGAGTTCGGGGCCCGCGGATGGCGAAGGGGATGCGCAAAAGCTCCGTGCTCTTT 1080

QY 1081 GTGGAGCGCACCACTTACCCCGAGGGGTAGCACTTCGCCCGCCCGCCAGGGAGGACCACT 1140
DB 1081 GTGGAGCGCACCACTTACCCCGAGGGGTAGCACTTCGCCCGCCCGCCAGGGAGGACCACT 1140

QY 1141 GCGAGCATTCACCGAGTGGTCCCTCAAGACAGAGCGGGGCGCAGCGGAGAGCTCCCA 1200
DB 1141 GCGAGCATTCACCGAGTGGTCCCTCAAGACAGAGCGGGGCGCAGCGGAGAGCTCCCA 1200

QY 1201 GAAGTATCTATGGGTGGGTTACTCTTCAAGCGCCCTGTAAGAGGGGACTTCCCATTC 1260
DB 1201 GAAGTATCTATGGGTGGGTTACTCTTCAAGCGCCCTGTAAGAGGGGACTTCCCATTC 1260

QY 1261 CCTCAGCCACCTGATGTCCCGACGACTGGTTCATCCAGCCACCAAGTGCCTCCCAACAA 1320
DB 1261 CCTCAGCCACCTGATGTCCCGACGACTGGTTCATCCAGCCACCAAGTGCCTCCCAACAA 1320

```

FEATURES		Helix Research Institute (JP)	
source	Location/Qualifiers		
	1..1811		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	659..1468		
CDS	/notes="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC32398.1"		
	/db_xref="GI:13121983"		
	/db_xref="RENTREMBL:CAC32398"		
	/translation="MKTPVELAVSGMOTLGLQHRGGRVVRKARSYVDETLFGSPAG TRTPDPFPWVEKARTGKGVKEASKALGAKSCETTPSRGSTFTLTFRKXNYRP ISHTPSYCDSELSFGSSEGAFCAPMAKGDAAKLRAIWLTPPTFRGSHSPRPREP LRAIHGAPSKTPPGKADSOGLSMGLHSSRLKRLGSLHSLHNPSTGHGATSAP HTNGPDLPSTSGVTFRSPPLVTSRARSVSISVPSPRGGATQKPKPWK"		
ORIGIN			
Query Match		100.0%; Score 1811; DB 6; Length 1811;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1811; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	AGTGCCCTGCGGCGCTCTGCGGCGCTAGTACACACGACCTGAGTGAGTGGCCACGAGGAC	60
Db	1	AGTGCCCTGCGGCGCTCTGCGGCGCTAGTACACACGACCTGAGTGAGTGGCCACGAGGAC	60
Qy	61	CCTCTCCATGTTTAGGACCTCTCTGGGCTTCAGAGCGTGGCGCCCGCCCTGGCGGAC	120
Db	61	CCTCTCCATGTTTAGGACCTCTCTGGGCTTCAGAGCGTGGCGCCCGCCCTGGCGGAC	120
Qy	121	TCCCCCATCGCGGGCGCAATGCTCCGGTCCGCTCCGAGTGGTGGTGGTGGTGGTGGT	180
Db	121	TCCCCCATCGCGGGCGCAATGCTCCGGTCCGCTCCGAGTGGTGGTGGTGGTGGTGGT	180
Qy	181	TGTTGCTGGGTGCAAGTCTGGGTTCTGGGTTCTGGATTCGCGGGCGGTTTACACGT	240
Db	181	TGTTGCTGGGTGCAAGTCTGGGTTCTGGGTTCTGGATTCGCGGGCGGTTTACACGT	240
Qy	241	AGCTGTGCGGCTCTCGGGTGGTCCGCGGCGGCTGCGGCGGCGGCTGCGGCGGCTAGGC	300
Db	241	AGCTGTGCGGCTCTCGGGTGGTCCGCGGCGGCTGCGGCGGCGGCTGCGGCGGCTAGGC	300
Qy	301	TGCGGGGCTCCGGGGCGCCAGGATTCGCGGCTGCAGATTGACGGGGATCCCGGATGCA	360
Db	301	TGCGGGGCTCCGGGGCGCCAGGATTCGCGGCTGCAGATTGACGGGGATCCCGGATGCA	360
Qy	361	CCGCGCGCCCGCGGCTCCTCAGCGAGCGGTCCAGCTGCTGGTGGGAAGAGTCCGGGAC	420
Db	361	CCGCGCGCCCGCGGCTCCTCAGCGAGCGGTCCAGCTGCTGGTGGGAAGAGTCCGGGAC	420
Qy	421	GGGTCCCTGAGGATCCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA	480
Db	421	GGGTCCCTGAGGATCCCGATGCTTACGAGCAAGATGCTCAGCTTTATAGGTGTGACCTA	480
Qy	481	CACATGTGACTTACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGCTACTTCA	540
Db	481	CACATGTGACTTACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGCTACTTCA	540
Qy	541	CAGTGTCTGTGAGAGGATTAATGAACAATGCTTGTAAAGCTTTTGCAGGAGGAGCC	600
Db	541	CAGTGTCTGTGAGAGGATTAATGAACAATGCTTGTAAAGCTTTTGCAGGAGGAGCC	600
Qy	601	TCGGAAGCAGGCGCTGCGCGAGACACACCTGCTCTCACAGGACACCAAGGAGGAGCC	660
Db	601	TCGGAAGCAGGCGCTGCGCGAGACACACCTGCTCTCACAGGACACCAAGGAGGAGCC	660
Qy	661	GAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGACAGCCCTGCGCCTTTCAGCACCGCTG	720
Db	661	GAAGACCCCGTGGAGCTGGCGCTCAGTGGGATGACAGCCCTGCGCCTTTCAGCACCGCTG	720
Qy	721	CCGAGGTGGCTACCGGCTCAAGGCCAGGACGCTCATATGTGATGAGACTCTGTTGGCAG	780
Db	721	CCGAGGTGGCTACCGGCTCAAGGCCAGGACGCTCATATGTGATGAGACTCTGTTGGCAG	780
Qy	781	CCCAGCAGGACCCCGGCTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAGGCTAA	840
Db	781	CCCAGCAGGACCCCGGCTACCCACCGGACTTCGATCCGCGCTTGGGTGGAGAGGCTAA	840
Qy	841	CAGAACAGAGGCTGGGCGAAGAGGAGCATCGAAGGCTTGGGGCGCAAGGGGAGCTGTGA	900
Db	841	CAGAACAGAGGCTGGGCGAAGAGGAGCATCGAAGGCTTGGGGCGCAAGGGGAGCTGTGA	900
Qy	901	GACACACCCCTCAAGGGGAGCACCACCCCTCACACCAAGGAAGAAACAATAACAG	960
Db	901	GACACACCCCTCAAGGGGAGCACCACCCCTCACACCAAGGAAGAAACAATAACAG	960
Qy	961	ACCCATCAGCACACCCCGCTTACTGTGATGAGTCCGTGTTGGCTCCCGATCTGAAGG	1020
Db	961	ACCCATCAGCACACCCCGCTTACTGTGATGAGTCCGTGTTGGCTCCCGATCTGAAGG	1020
Qy	1021	CGCAGCTTCCGGGCGCGGATGGCGAAGGGGATGCCGCAAGCTCCGTGCTCTCTT	1080
Db	1021	CGCAGCTTCCGGGCGCGGATGGCGAAGGGGATGCCGCAAGCTCCGTGCTCTCTT	1080
Qy	1081	GTGACGCGCACACCTTACCCCGAGGGTAGCCACTCCGCGCGCCCGAGGGAGGCCACT	1140
Db	1081	GTGACGCGCACACCTTACCCCGAGGGTAGCCACTCCGCGCGCCCGAGGGAGGCCACT	1140
Qy	1141	GGGAGCATTCACCCAGCTGCTCCCTCCAGACAGAGCGGGGCGAGGCGAGTCCCA	1200
Db	1141	GGGAGCATTCACCCAGCTGCTCCCTCCAGACAGAGCGGGGCGAGGCGAGTCCCA	1200
Qy	1201	GAAATTAATCTATGGTGGGTTACACTCTTTCAGCGCCCTTGAAGCGGGGACTTTCCCAATC	1260
Db	1201	GAAATTAATCTATGGTGGGTTACACTCTTTCAGCGCCCTTGAAGCGGGGACTTTCCCAATC	1260
Qy	1261	CCTCACCCACCTGAATGTCCCAGACCTGCTGTCATCCAGCACCAGTGGCCCCCACA	1320
Db	1261	CCTCACCCACCTGAATGTCCCAGACCTGCTGTCATCCAGCACCAGTGGCCCCCACA	1320
Qy	1321	TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACCTTCCGAGGCGCCCTGGTGAC	1380
Db	1321	TGGGCTCAGGATCTCAGGCTTCCAGCTCAGGGGTGACCTTCCGAGGCGCCCTGGTGAC	1380
Qy	1381	TTCCAGGCTCAGTTCAGTTAGCTTTCAGTGCCATCTACCCAGAGGAGTGGGGCCAC	1440
Db	1381	TTCCAGGCTCAGTTCAGTTAGCTTTCAGTGCCATCTACCCAGAGGAGTGGGGCCAC	1440
Qy	1441	CCAGAAACCAAGCCCTTGGAAATGATCTCTTTCATCAGGTTGCTATGGGGCCAC	1500
Db	1441	CCAGAAACCAAGCCCTTGGAAATGATCTCTTTCATCAGGTTGCTATGGGGCCAC	1500
Qy	1501	GGCCACAGGATGGCCCTTGGCGGTTAGGAGGACATTCATCACCAGGGAACCCACAGG	1560
Db	1501	GGCCACAGGATGGCCCTTGGCGGTTAGGAGGACATTCATCACCAGGGAACCCACAGG	1560
Qy	1561	TATTAAGAGGCTGTGGGGGAGACAGACATAGAGGGGTGGGAGTGCCTCCCTTT	1620
Db	1561	TATTAAGAGGCTGTGGGGGAGACAGACATAGAGGGGTGGGAGTGCCTCCCTTT	1620
Qy	1621	ATCCCTGACATCTCTAGTTCGATTTTCTCCCGATTCGCGATTTGGGGGCGAC	1680
Db	1621	ATCCCTGACATCTCTAGTTCGATTTTCTCCCGATTCGCGATTTGGGGGCGAC	1680
Qy	1681	CTCTAAGATGCTCTCTCAGCCCTGTCTCAACATCTCCAAATTAGTCCAAACCCAGG	1740
Db	1681	CTCTAAGATGCTCTCTCAGCCCTGTCTCAACATCTCCAAATTAGTCCAAACCCAGG	1740
Qy	1741	GGCCTGGGACCTCCACATCTCCTTGTGTCGCAAGTGGCAATTAACGGGCTGAT	1800
Db	1741	GGCCTGGGACCTCCACATCTCCTTGTGTCGCAAGTGGCAATTAACGGGCTGAT	1800
Qy	1801	TGCCAACCTGG 1811	
Db	1801	TGCCAACCTGG 1811	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 03:23:33 ; Search time 7205 Seconds
(without alignments)
10894.411 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgcctgcggccctcggcg.....cggcgtgatgcgaacctgg 1811

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rod:*
36: em.htg.man:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1811	100.0	1811	6	AX077432	Sequence
2	1811	100.0	1811	6	AX075358	Homo sapi
3	1631.4	90.1	1839	9	AY007158	Homo sapi
4	1621	89.5	1775	6	AX880378	Sequence
5	1621	89.5	1775	6	BD158346	Primer fo
6	1621	89.5	1775	9	AK027733	Homo sapi
7	1597.6	88.2	1709	9	AK123762	Homo sapi
8	1434.8	79.2	1614	6	AX880429	Sequence
9	1434.8	79.2	1614	6	BD158376	Primer fo
10	1434.8	79.2	1614	9	AK027741	Homo sapi
11	1371	75.7	1407	9	BC022092	Homo sapi
12	854	47.2	199522	9	AC089899	Homo sapi
13	661.4	36.5	784	6	AX869030	Sequence
14	661.4	36.5	784	6	BD149092	Primer fo
15	562	31.0	597	6	AX869102	Sequence
16	562	31.0	597	6	BD149164	Primer fo
17	550	30.4	816	6	AX781130	Sequence
18	494	27.3	561	6	AX874097	Sequence
19	494	27.3	561	6	BD154159	Primer fo
20	472.8	26.1	1485	10	BC021365	Mus muscu
21	369	20.4	572	6	AX779761	Sequence
22	353.8	19.5	263827	2	AC094579	Rattus no
23	353.8	19.5	274026	2	AC129633	Rattus no
24	341.8	18.9	578	6	AX874033	Sequence
25	341.8	18.9	578	6	BD154095	Primer fo
26	299.2	16.5	237928	2	AC137208	Rattus no
27	273.2	15.1	221622	10	AC125183	Mus muscu
28	241	13.3	199165	2	AC131195	Rattus no
29	186.2	10.3	60319	2	AC102037	Mus muscu
30	130.8	7.2	292	6	AX315440	Sequence
31	126	7.0	60319	2	AC102037	Mus muscu
32	125	6.9	232481	2	AC121112	Mus muscu
33	82.8	4.6	125020	9	AF429315	Homo sapi
34	61.6	3.4	125020	9	AF429315	Homo sapi
35	54.4	3.0	39970	9	AP001167	Homo sapi
36	54.4	3.0	146451	2	AC019309	Homo sapi
37	54.4	3.0	185984	9	AP000942	Homo sapi
38	54.2	3.0	146574	2	AC023459	Homo sapi
39	54.2	3.0	168860	9	AL160175	Human DNA
40	54	3.0	2000	6	AX655393	Sequence
41	53.8	3.0	3116	9	HSAC001552	Homo sapi
42	53.8	3.0	84544	2	AC009012	Homo sapi
43	53.8	3.0	159420	2	AC009017	Homo sapi
44	53.8	3.0	160672	9	AC008608	Homo sapi
45	53.6	3.0	89992	9	AL135794	Human DNA

ALIGNMENTS

RESULT 1
AX077432
LOCUS AX077432 1811 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 3 from Patent WO0107607.
ACCESSION AX077432
VERSION AX077432.1 GI:13121982
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota, T., Isogai, T., Nishikawa, T. and Kawai, Y.
Full length cdna clones and proteins encoded thereby
Patent: WO 0107607-A 3 01-FEB-2001;
JOURNAL

This Page Blank (uspto)

```

Db 444 TCACTAAGACACAGTTCTTTATCTGTAAATAGTGACATGAATAGCATCTTGAAGGGC 385
QY 547 TGTTCAGAGGATTAAATGAACAATGCTTTGTAAAGCTCTTTGCAGGAGGAGCCTCGAA 606
Db 384 TACTGTGAGGATTATATGAATATGAACGCAATGTACTTAGCAGGTTCTGGACTCAAA 325
QY 607 GCAGGCGCTGCCCGCAGAG 626
Db 324 AGGCAACTGTTATGATGAG 305

```

RESULT 14

```

US-10-292-798-1243
; Sequence 1243, Application US/10292798.
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1243
; LENGTH: 43981
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(43981)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(470)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12636)..(12748)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24946)..(25088)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32801)..(32997)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (41097)..(41379)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (43585)..(43781)
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (2811)..(2910)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6914)..(6914)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6923)..(6923)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6932)..(6932)
; OTHER INFORMATION: a, t, c, g, unknown or other

```

```

; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6956)..(6956)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6959)..(6960)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (6964)..(6964)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (7905)..(7905)
; OTHER INFORMATION: a, t, c, g, unknown or other
; US-10-292-798-1243

```

```

Query Match 2.8%; Score 50.4; DB 16; Length 43981;
Best Local Similarity 71.7%; Pred. No. 0.00061;
Matches 66; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 495 CTTACCTCAGTTTGTGATCCGTTAAATGGAACAATTCGAAGCTACTTCCACAGTCTGTGTGAGA 554
Db 18915 CCTCAGTTCTCTTATCTGTTAGTGGAGATAATAACAATACTACCTCAAAGGGTTGTAGGA 18974
QY 555 GGATTAAATGAACAATGCTTGTAAAGCTCTT 586
Db 18975 GGATTAAATGAATTAATATTCGTAAGCACIT 19006

```

RESULT 15

```

US-09-867-701-7274
; Sequence 7274, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7274
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-867-701-7274

```

```

Query Match 2.8%; Score 50; DB 9; Length 316;
Best Local Similarity 67.0%; Pred. No. 0.00015;
Matches 71; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

```

QY 490 CTTACCTCAGTTTGTGATCCGTTAAATGGAACAATTCGAAGCTACTTCCACAGTCTGTGT 549
Db 7 CCTTACATCAATCTCCTTATCTGTGAAATGGGAATAATAAAATCTACCTCAGAAATTAT 66
QY 550 TGAGAGGATTAATGAACAATGCTTGTAAAGCTCTTTGCAGGAGG 595
Db 67 TAACAGATTAATGAGATAATCCATGGAAGCTAGTAGCAGCATG 112

```

```

Search completed: August 17, 2004, 09:56:45
Job time : 860 secs

```



```
; Sequence 288030, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match      2.9%; Score 52.8; DB 16; Length 479;
Best Local Similarity 61.8%; Pred. No. 2.5e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 474 TGACCTACACATGACTTCACCTCAGTTTGTGATCCGTAATGACAAATTCGAGC 533
DB 89 TGGCTGGGGAGGTAACCTTCACTCAGTTTCTCATCTGACCACTGGGTACAGTTGGC 148

QY 534 TACTTCACAGTGTGTTGAGAGGATTAATGACAAATGCTTGTAAAGCTCTTTGCAGGA 593
DB 149 TGATTCAAGAGTGGTTGTAAGATTAACATGACATGACGTGTAGAAAGCACTTAGCAGAA 208

QY 594 GGGAGCCTCGGAACA 609
DB 209 TGCCTAGTCCAAAGCA 224

RESULT 9
US-10-242-355-1101
; Sequence 1101, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242.355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATGACAAATTCGAAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATATAAATCTACCTCAGAAATAT 94

QY 550 TGAGAGGATTAAATGAAACAATGCTTGTAAAGCTCTTTCGAGGAGG 595
DB 95 TAAACAAATTAATGAGATATCCATGGAAAGCTAGTAGCAGCATG 140
```

```
Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATGACAAATTCGAAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATATAAATCTACCTCAGAAATAT 94

QY 550 TGAGAGGATTAAATGAAACAATGCTTGTAAAGCTCTTTCGAGGAGG 595
DB 95 TAAACAAATTAATGAGATATCCATGGAAAGCTAGTAGCAGCATG 140

RESULT 10
US-10-242-355-1102
; Sequence 1102, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242.355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1102
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-355-1102

Query Match      2.8%; Score 51.6; DB 16; Length 680;
Best Local Similarity 67.9%; Pred. No. 6.5e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 490 CTTACCTCAGTTTGTGATCCGTAATGACAAATTCGAAAGCTACTTACAGTGTGT 549
DB 35 CCTTACCTCAATCTCCTTATCTGTGAAATGGGAATATAAATCTACCTCAGAAATAT 94
```

Db 339 GTGGCTACCGGGTCAAGGCCAGACGTCATATGTGATGAGACTCTGTTTGGCAGCCAG 398
Qy 786 CAGGACCCCGGCTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGTTAAACAGAA 845
Db 399 CAGGACCCCGGCTACCCACCGGACTTCGATCCGCCCTGGTGGAGAGGTTAAACAGAA 458
Qy 846 CCAGAGCGGTGGCAGGAGGACATCG 871
Db 459 CCAGAGCGGTGGCAGGAGGACATCG 484

RESULT 5
US-09-864-408A-8425
; Sequence 8425, Application US/09864408A
; Publication No. US2004000947A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US2004000947A1e1 Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8425
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-8425

Query Match 7.2%; Score 130.8; DB 11; Length 292;
Best Local Similarity 72.8%; Pred. No. 4.9e-29;
Matches 182; Conservative 0; Mismatches 67; Indels 1; Gaps 1;
Qy 639 CACGAGGACCCAGGACGATGAGACCCCGGTGG-AGCTGGCCGTCAGTGGGATGCAG 697
Db 43 CACCACACCCACTGGGCAACGTGAGACCCCGGTGGCAGCGGCCAGCAGCGGATGAA 102
Qy 698 ACCTCGGCTTCAGCACCGCTGCCAGAGTGGCTACCGGGTCAAGGCCAGGACGTCATAT 757
Db 103 GCCATCCACTTCAGCACCGCAGTCCACCAGCTACAGGGTCAAGGCCAGGACATCCTAT 162
Qy 758 GTGGATGAGACTCTGTTGGCAGCCAGCAGGACCCGCCCTACCCACCGGACTTCGAT 817
Db 163 GTGGATGAGACCTCTGTTGGCAGCCAGCAGGACCCAGCAGGACCCAGCAGCTTGAC 222
Qy 818 CCGCCCTGGGTGAGAGGCTAACAGAACAGGCGGTGGGCAAGAGGACATCGAAGGCC 877
Db 223 CCACCTGGGTGAGAGGCTAACAGAACAGGCGGTGGGCAAGAGGACATCGAAGGCC 282
Qy 878 TTGGGGGCGA 887
Db 283 TCTCTGGCCA 292

RESULT 6
US-09-908-975-4892
; Sequence 4892, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; FILE REFERENCE: THAT POPULATE A TRANSCRIPTOME
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4892
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-4892

Query Match 3.3%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 8.2e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 164 TGTGCTGGGTGCTCCCTGCTGTGTGCTGGTGCAAAAGTCTCGGGTTCCTGGATTC 223
Db 1 TGTGCTGGGTGCTCCCTGCTGTGTGCTGGTGCAAAAGTCTCGGGTTCCTGGATTC 60

RESULT 7
US-10-027-632-288030
; Sequence 288030, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 288030
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288030

Query Match 2.9%; Score 52.8; DB 13; Length 479;
Best Local Similarity 61.8%; Pred. No. 2.5e-05;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 474 TGACCTACACATGTGACTTCACCTCAGTTTGTGATCCGTAAATGGACAAATTCGAAGC 533
Db 89 TGGCTGGGAGGTAACTTCACTCAGTGTCTCACTGACCATGGGTACAGTTGCC 148
Qy 534 TACTTCACAGTGTGTTGAGAGGATTAATGAACAAATGCTTGAAGCTCTTTGAGGA 593
Db 149 TGATTCAAGGTGGTTGTAAAGAAATTAACGACATGCGTGTGAAGAGCACTTAGCAGAA 208
Qy 594 GGGAGCCTCGGAAGCA 609
Db 209 TGCCTAGTCCAAGCA 224
RESULT 8
US-10-027-632-288030

QY 922 CACCCACCCCTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACACACCCGTC 981
 Db 254 CACCCACCCCTCACACCAAGGAAGAAACAAATACAGACCCATCAGCCACACACCCGTC 313
 QY 982 TTAATGTGATGAGTCGCTGTTGGCTCCGATCTGAAGCGCCGAGCTTGGGGCCCGCG 1041
 Db 314 TTAATGTGATGAGTCGCTGTTGGCTCCGATCTGAAGCGCCGAGCTTGGGGCCCGCG 373
 QY 1042 GATGGCAAGGGGATCGCCCAAGCTCCGTGCTCTTGTGACGCCACCACTACCC 1101
 Db 374 GATGGCAAGGGGATCGCCCAAGCTCCGTGCTCTTGTGACGCCACCACTACCC 433
 QY 1102 CAGGGGTAGCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGG 1161
 Db 434 CAGGGGTAGCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGG 493
 QY 1162 TCCCTCCAGACAGAGCCGGGCGCCAGGCGAGACTCCAGAACTTCTATGGTGGGTT 1221
 Db 494 TCCCTCCAGACAGAGCCGGGCGCCAGGCGAGACTCCAGAACTTCTATGGTGGGTT 553
 QY 1222 ACACCTCTTACGCCCCCTGAAGGGGGACTTTCCCATTTCCCTCACCCACCTGAATGTCC 1281
 Db 554 ACACCTCTTACGCCCCCTGAAGGGGGACTTTCCCATTTCCCTCACCCACCTGAATGTCC 613
 QY 1282 CAGCACTGTGATCAGCCACCAAGTGCCTCCCGCCACCAATATGGCCCTCAGGATCTCAGGCC 1341
 Db 614 CAGCACTGTGATCAGCCACCAAGTGCCTCCCGCCACCAATATGGCCCTCAGGATCTCAGGCC 673
 QY 1342 TTCCACCTCAGGGGTGACCTTCGGAGCCCGCTGGT 1377
 Db 674 TTNCAGCTCAGGGGTGACCTTCGGAGCCCGCTGGT 709
 *
 RESULT 3
 US-09-918-995-1667
 ; Sequence 1667, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1667
 ; LENGTH: 477
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(477)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-1667
 Query Match 24.1%; Score 436.6; DB 10; Length 477;
 Best Local Similarity 99.1%; Pred. No. 1.3e-121;
 Matches 439; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 986 TGTGATGAGTCGCTGTTGGCTCCGATCTGAAGGGCCAGCTTCGGGGCCCGCGGATG 1045
 Db 35 TGGAAATTCGTCGCTGTTGGCTCCGATCTGAAGGGCCAGCTTCGGGGCCCGCGGATG 94
 QY 1046 GCGAAGGGGATCGCGAAAGCTCCGTGCTCTCTGTGACGCCACCACTACCCCGAG 1105
 Db 95 GCGAAGGGGATCGCGAAAGCTCCGTGCTCTCTGTGACGCCACCACTACCCCGAG 154
 QY 1106 GGTAGCACTCGCCCGCCCGCCAGGAGGACCACTGCGAGCCATTCACCCAGCTGTCC 1165

Db 155 GGTAGCCACTCGCCCGCCCGCCAGGGAGGACCACTCGAGGCCATTACCCAGCTGGTCCC 214
 QY 1166 TCCAGACAGAGCCGGGCGCCAGCGCAGACTCCCGAAGTTATCTATGGGTGGTTACAC 1225
 Db 215 TCCAGACAGAGCCGGGCGCCAGCGCAGACTCCCGAAGTTATCTATGGGTGGTTACAC 274
 QY 1226 TCTTACGCCCCCTTGAAGCGGGGACTTTCCGATTTCCCTCACCCACCTGAATGTCCCCAGC 1285
 Db 275 TCTTACGCCCCCTTGAAGCGGGGACTTTCCGATTTCCCTCACCCACCTGAATGTCCCCAGC 334
 QY 1286 ACTGTCATTCAGCACCCAGTGCCTCCCGCCACACAAATGGGCTCAGGATCTCAGGCTTCC 1345
 Db 335 ACTGTCATTCAGCACCCAGTGCCTCCCGCCACACAAATGGGCTCAGGATCTCAGGCTTCC 394
 QY 1346 ACCTCAGGGGTGACCTTCGCGAGCCCTCGTGTGATCTTCAGGGGCTCGCTCAGTTAGCAT 1405
 Db 395 ACCTCAGGGGTGACCTTCGCGAGCCCTCGTGTGATCTTCAGGGGCTCGCTCAGTTAGCAT 454
 QY 1406 TCAGTGCCATCTACCCACGAGC 1428
 Db 455 TCAGTGCCATCTACCCACGAGC 477
 RESULT 4
 US-09-918-995-31166
 ; Sequence 31166, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 31166
 ; LENGTH: 484
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(484)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-31166
 Query Match 22.7%; Score 410.6; DB 10; Length 484;
 Best Local Similarity 93.9%; Pred. No. 9.5e-114;
 Matches 419; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 426 CCTGAGGATCCCGATGCTTACGAGCCAGAGTCTCAGCTTTATAGGTGTGACCTACACAT 485
 Db 39 CNGNAANNNNNNNAGNANAGCCAGAGTCTCAGCTTTATAGGTGTGACCTACTCAT 98
 QY 486 GTGACTTCACCTCAGTTTGTGATCCCGTAAATGCAAAATTCGAAGTACTTCACAGTG 545
 Db 99 GTGACTTTACCTCAGTTTGTGATCCCAAAATGCAAAAGGCGAAGTACTTCACAGAG 158
 QY 546 CTGTTGAGAGATTAAATGAACAAATGCTTGAAGCTCTTTGAGGAGGAGGCTCGGA 605
 Db 159 CTGTTGAGAGATTAAATGAACAAATGCTTGAAGCTCTTTGAGGAGGAGGCTCGGA 218
 QY 606 AGCAGGGCTCGCGGCGAGACACACCTGCTGTCCAGGAGCCACAGGAGCATGAAGA 665
 Db 219 AGCAGGGCTCGCGGCGAAGCAACCTGCTGTCCAGGAGCCACAGGAGCATGAAGA 278
 QY 666 CCCCCTGAGAGCTGGCGCTCAGTGGGATGACAGACCTTCGGCTTCAGCAGCGCTGCCAG 725
 Db 279 CCCCCTGAGAGCTGGCGCTCAGTGGGATGACAGACCTTCGGCTTCAGCAGCGCTGCCAG 338
 QY 726 GTGGCTACCGGGTCAAGGCCAGGAGCTCATATGTGGATGAGACTCTGTTGGCAGCCAG 785

Db 130 CCCAGGCAATTCGGGCTGCAGATTGACGGGGATCCGGATGCACCGCGGCCCCCGCGC 189
Qy 377 CTTACCGAGCGGTCCAGACTCTGGTGGGAAGAGGTGCGGGACGGGTCCCTGAGATCC 436
Db 190 CCTCACCGAGCGGTCCAGACTCTGGTGGGAAGAGGTGCGGGACGGGTCCCTGAGATCC 249
Qy 437 CGATGCTCAGACGCAAGATGCTCAGCTTTATAGGTGTGACCTTACACATGTGACTTCACC 496
Db 250 CGATGCTCAGACGCAAGATGCTCAGCTTTATAGGTGTGACCTTACACATGTGACTTCACC 309
Qy 497 TCAGTTTGTGATCCGTAAATCGACAAATTCGAAGCTACTTCAACAGTCTGTGTGAGAG 556
Db 310 TCAGTTTGTGATCCGTAAATCGACAAATTCGAAGCTACTTCAACAGTCTGTGTGAGAG 369
Qy 557 ATTAATGAACAAATCTGTTGAAGCTCTTTCAGAGGAGGAGCTCGGAAGCAGGCGCTG 616
Db 370 ATTAATGAACAAATCTGTTGAAGCTCTTTCAGAGGAGGAGCTCGGAAGCAGGCGCTG 429
Qy 617 GCGGCGAGACACACCTGCTGTCCACGAGGACACACAGGAGCATGAAGACCCCGGTGGAG 676
Db 430 GCGGCGAGACACACCTGCTGTCCACGAGGACACACAGGAGCATGAAGACCCCGGTGGAG 489
Qy 677 CTGGCGGTGAGTGGGATGAGACCTCTGGCTTCAGCAGCGCTCGCGAGGTGGCTACCGG 736
Db 490 CTGGCGGTGAGTGGGATGAGACCTCTGGCTTCAGCAGCGCTCGCGAGGTGGCTACCGG 549
Qy 737 GTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTTGGAGCCCGACGACCCCGG 796
Db 550 GTCAAGGCCAGGACGTCATATGTGATGAGACTCTGTTGGAGCCCGACGACCCCGG 609
Qy 797 CTTACCCACCGGACTTCGATCCGCTCGGCTGGAGAGGCTAAAGAACCCCGGTGGAG 856
Db 610 CTTACCCACCGGACTTCGATCCGCTCGGCTGGAGAGGCTAAAGAACCCCGGTGGAG 669
Qy 857 GGCAAGGAGCATCGAGGCTTGGGGCAAGAGGAGGTGTGAGACCCCGGTGGAG 916
Db 670 GGCAAGGAGCATCGAGGCTTGGGGCAAGAGGAGGTGTGAGACCCCGGTGGAG 729
Qy 917 GGCAGCACCCCACTTCACCAAGGAAGAAACAAATACAGACCCATCAGCCACACC 976
Db 730 GGCAGCACCCCACTTCACCAAGGAAGAAACAAATACAGACCCATCAGCCACACC 789
Qy 977 CCGCTTACTGATGATGCTGTTGGCTCCGATCTGAAGGCGGAGCTTCGGGCG 1036
Db 790 CCGCTTACTGATGATGCTGTTGGCTCCGATCTGAAGGCGGAGCTTCGGGCG 849
Qy 1037 CCGCGGATGCGGAGGGGATGCGGCAAGAGCTCCGTGCTCTCTTGTGAGCGCACACCT 1096
Db 850 CCGCGGATGCGGAGGGGATGCGGCAAGAGCTCCGTGCTCTCTTGTGAGCGCACACCT 909
Qy 1097 ACCCCAGGGGTAGCACTCGCCCGCCAGGGAGGACCACTGCGAGCCATTCACCCA 1156
Db 910 ACCCCAGGGGTAGCACTCGCCCGCCAGGGAGGACCACTGCGAGCCATTCACCCA 969
Qy 1157 GCTGGTCCCTTCAAGACAGAGCGGGCCAGCGGAGACTCCAGAGGTTATCTATGGT 1216
Db 970 GCTGGTCCCTTCAAGACAGAGCGGGCCAGCGGAGACTCCAGAGGTTATCTATGGT 1029
Qy 1217 GGGTTACATCTTCAAGCGCCCTTGAAGCGGGACTTTCCCATTCCTTCCACCCACCTGAAT 1276
Db 1030 GGGTTACATCTTCAAGCGCCCTTGAAGCGGGACTTTCCCATTCCTTCCACCCACCTGAAT 1089
Qy 1277 GTCCCGAGCACTGCTATCCAGCCACAGTGTGCCCCCAGCAAAATGGGCTCAGGATCTC 1336
Db 1090 GTCCCGAGCACTGCTATCCAGCCACAGTGTGCCCCCAGCAAAATGGGCTCAGGATCTC 1149
Qy 1337 AGGCTTCCACGTCAGGGGTGACCTTCCGAGGCCCCCTGGTGAATTCAGAGGCTCGCTCA 1396
Db 1150 AGGCTTCCACGTCAGGGGTGACCTTCCGAGGCCCCCTGGTGAATTCAGAGGCTCGCTCA 1209
Qy 1397 GTTAGCATTTCAATGCTATCCCAAGCAGGTTGGGGCCACCCAGAAACCAAGGCC 1456
Db 1210 GTTAGCATTTCAATGCTATCCCAAGCAGGTTGGGGCCACCCAGAAACCAAGGCC 1269

RESULT 2

US-10-452-858C-78
; Sequence 78, Application US/10452858C
; Publication No. US20040086945A1
; GENERAL INFORMATION:
; APPLICANT: Sreekrishna, Kotikanyadanam
; APPLICANT: Gerwe, Gina S.
; APPLICANT: Toerner, Daniel R.
; TITLE OF INVENTION: HAIRLESS PROTEIN-INTERACTING PARTNER COMPLEXES AND METHODS THEREOF
; FILE REFERENCE: 8956P
; CURRENT APPLICATION NUMBER: US/10/452,858C
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 709
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)..(707)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (670)..(695)
; OTHER INFORMATION: n represents a, c t or g
US-10-452-858C-78

Query Match 38.0%; Score 688.2; DB 17; Length 709;
Best Local Similarity 99.1%; Pred. No. 8.2e-198;
Matches 690; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 682 CGTCAGTGGGATGAGACCCCTCGGCTTCAGACACCGCTGCGAGGTGGTACCGGGTCAA 741
Db 14 CGTCAGTGGGATGAGACCCCTCGGCTTCAGACACCGCTGCGAGGTGGTACCGGGTCAA 73
Qy 742 GGCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGAGGACCCGGCTAC 801
Db 74 GGCAGGACGTCATATGTGATGAGACTCTGTTGGCAGCCCGACGAGGACCCGGCTAC 133
Qy 802 CCCACCGACTTCGATCCGCTCGGCTGGGAGAGAGGCTTAAACAGAACCCAGAGCGGTGGGCAA 861
Db 134 CCCACCGACTTCGATCCGCTCGGCTGGGAGAGAGGCTTAAACAGAACCCAGAGCGGTGGGCAA 193
Qy 862 GGAGGATCGAAGCGCTTGGGGGCAAGGGGAGGCTGTGAGACACCCCTCAAGGGGCGAG 921
Db 194 GGAGGATCGAAGCGCTTGGGGGCAAGGGGAGGCTGTGAGACACCCCTCAAGGGGCGAG 253

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 05:25:54 ; Search time 856 Seconds

(without alignments)
10380.685 Million cell updates/sec

Title: US-10-031-589-3

Perfect score: 1811

Sequence: 1 agtgcctggcgcctcgcg.....cggcggtattgccaaactgg 1811

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1610.2	88.9	1826	13	US-10-276-774-361
2	688.2	38.0	709	17	US-10-452-858C-78
3	436.6	24.1	477	10	US-09-918-995-1667
4	410.6	22.7	484	10	US-09-918-995-31166
5	130.8	7.2	292	11	US-09-864-408A-8425
6	60	3.3	60	10	US-09-908-978-4892
7	52.8	2.9	479	13	US-10-027-632-288030
8	52.8	2.9	479	16	US-10-027-632-288030
9	51.6	2.8	680	16	US-10-242-355-1101
10	51.6	2.8	680	16	US-10-242-355-1102
11	50.8	2.8	343	15	US-10-242-355-240
12	50.4	2.8	741	13	US-10-027-632-102308
13	50.4	2.8	741	16	US-10-027-632-102308
14	50.4	2.8	43981	16	US-10-292-798-1243

15	50	2.8	316	9	US-09-867-701-7274	Sequence 7274, Ap
16	49.8	2.7	51657	16	US-10-057-4758-10475	Sequence 10475, A
17	49.8	2.7	51657	16	US-10-154-884B-10475	Sequence 10475, A
18	49.8	2.7	174493	9	US-09-804-471A-3	Sequence 3, Appli
19	49.8	2.7	174493	15	US-10-238-709-3	Sequence 3, Appli
20	49.8	2.7	174493	17	US-10-724-594-3	Sequence 7271, Ap
21	48.4	2.7	267	9	US-09-867-701-7271	Sequence 7264, Ap
22	48.4	2.7	304	9	US-09-867-701-7264	Sequence 4, Appli
23	47.4	2.6	24295	17	US-10-137-277A-4	Sequence 11834, A
24	46.6	2.6	382	15	US-10-138-846-11834	Sequence 11834, A
25	46.6	2.6	653	15	US-10-138-846-3495	Sequence 3495, Ap
26	46.6	2.6	677	13	US-10-027-632-139069	Sequence 139069, A
27	46.6	2.6	677	16	US-10-027-632-139069	Sequence 139069, A
28	45.8	2.5	172637	9	US-09-805-458A-3	Sequence 3, Appli
29	45.8	2.5	822900	16	US-10-292-798-1393	Sequence 1393, Ap
30	45.6	2.5	767	13	US-10-027-632-7958	Sequence 7958, Ap
31	45.6	2.5	767	15	US-10-027-632-7958	Sequence 7958, Ap
32	45.6	2.5	858	13	US-10-027-632-33700	Sequence 33700, A
33	45.6	2.5	858	16	US-10-027-632-33700	Sequence 33700, A
34	45.6	2.5	181684	13	US-10-087-192-790	Sequence 790, App
35	45.4	2.5	22756	10	US-09-764-891-6609	Sequence 6609, Ap
36	45.4	2.5	22756	15	US-10-091-572-473	Sequence 473, App
37	45.2	2.5	528	9	US-09-796-692-5239	Sequence 5239, Ap
38	45.2	2.5	528	15	US-10-040-862-5239	Sequence 5239, Ap
39	45.2	2.5	528	16	US-10-057-4758-5239	Sequence 5239, Ap
40	45.2	2.5	528	16	US-10-154-884B-5239	Sequence 5239, Ap
41	45.2	2.5	46143	13	US-10-087-192-640	Sequence 640, App
42	45	2.5	505	13	US-10-027-632-190302	Sequence 190302, A
43	45	2.5	505	16	US-10-027-632-190302	Sequence 190302, A
44	45	2.5	607	13	US-10-027-632-35453	Sequence 35453, A
45	45	2.5	607	13	US-10-027-632-35454	Sequence 35454, A

ALIGNMENTS

RESULT 1

US-10-276-774-361
; Sequence 361, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 361
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-361

Query Match	88.9%	Score	1610.2	DB	13	Length	1826
Best Local Similarity	99.8%	Pred. No.	0				
Matches	1612	Conservative	0	Mismatches	3	Indels	0
Gaps	0						
QY	197	AGTGTCTGGTTCGGTTCTTGGATTTCGGGGCGGTTCACACGTAGCTGTGCGGGTCC	256				
DB	10	ATTCCTGGTTCTGGTTCTTGGATTTCGGGGCGGTTCACACGTAGCTGTGCGGGTCC	69				
QY	257	TGCGGTGAGTCGTCGCGCGGTTCGCGGGAGCGGCTAGGTCGCGGGTTCGCGGG	316				
DB	70	TGCGGTGAGTCGTCGCGCGGTTCGCGGGAGCGGCTAGGTCGCGGGTTCGCGGG	129				
QY	317	CCCGAGGATTCGCGGTTCGAGTTGACGGGATCCCGGATGACCGCGCCCGCGGC	376				

This page Blank (uspio)